Investigation into Risk and Uncertainty: Identifying Coefficient of Variation Benchmarks for Air Force ACAT I Programs

THESIS

Shaun T. Carney, Captain, USAF

AFIT-ENV-13-M-05

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A

AFIT-ENV-13-M-05

Investigation into Risk and Uncertainty: Identifying Coefficient of Variation Benchmarks for Air Force ACAT I Programs

THESIS

Presented to the Faculty

Department of Systems Engineering and Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Cost Analysis

Shaun T. Carney, BS

Captain, USAF

March 2013

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION IS UNLIMITED

AFIT-ENV-13-M-05

Investigation into Risk and Uncertainty: Identifying Coefficient of Variation Benchmarks for Air Force ACAT I Programs

Shaun T. Carney, BS
Captain, USAF

Approved:

Jonathan D. Ritschel, Lt Col, USAF (Chairman)	Date
Edward D. White, Ph. D (Member)	Date
Dustin S. McGlothen (Member)	Date

AFIT-ENV-13-M-05

Abstract

The Air Force is searching for measures to reduce cost growth in defense acquisitions during times of constrained budgets across the Department of Defense. Previous DoD cost growth studies found typical cost growth in defense acquisitions is around forty-six to sixty percent of the original estimate. The research in this study addresses the identification of risk and uncertainty benchmarks by providing decision makers with coefficient of variation ranges for cost estimates. The hypothesis is that if cost estimates include enough risk and uncertainty adjustments then the DoD could more accurately estimate programs and therefore reduce cost growth. The intentions of the study are to recommend coefficient of variation (CV) ranges for Air Force Acquisition programs, determine if different CV ranges should be used based on platform type, and determine if CV decreases over the course of the program's acquisition lifecycle. This research is unique to previous cost growth studies because it employs source data from program offices in addition to Selective Acquisition Reports to answer the research questions. The analysis found that the Air Force should enhance the CV review process to ensure cost estimates have CVs between 41-74% during Milestone A, 31-54% during Milestone B, and 23-32% during Milestone C. It is recommended that Selective Acquisition Reports include the CV utilized to develop the current estimate. The analysis also found that CVs are analogous among platform types. There is not a necessity to operationalize CV ranges by product center or weapon system type. Lastly, the research found that CVs decrease as a program matures through the acquisition lifecycle.

AFIT-ENV-13-M-05

To my wife and children, for their love, support, and patience through this journey.

Acknowledgments

I would like to thank Dustin McGlothen and his staff, Mike Seibel and Charlie Clark, for searching through their archives tirelessly for program office data to make this thesis possible. I would like to thank Lt Col Ritschel for his guidance, wisdom, leadership, and encouragement to pursue this interesting area of research for the cost community. I would like to thank Dr. White for providing me the tools necessary to complete the analysis. I would like to thank my classmates for their support as we endured this process together. Lastly and most importantly, I would like to thank my wife and children for their support and understanding.

Shaun T. Carney

Table of Contents

Page

Page

List of Figures

List of Tables

List of Equations

Investigation into Risk and Uncertainty: Identifying Coefficient of Variation Benchmarks for Air Force ACAT I Programs

I. Introduction

The current economic climate necessitates that Department of Defense (DoD) leaders make better decisions allocating resources. The department invested $131B in procurement programs in 2011 (DoD, 2012: 2). The investment of such a large amount of taxpayer dollars in defense acquisitions requires accurate cost estimates to aid decision makers in allocating resources. Unfortunately, cost estimators are notorious for underestimating the procurement cost of new technologies and weapon systems. The total acquisition cost of DoDs 2007 portfolio of major defense acquisition programs has grown by nearly $300B over initial estimates (GAO, 2008:4). The significant cost growth has led to high visibility on cost estimates of major defense acquisition programs. The pressure to contain cost growth on DoD leaders compels them to demand more accurate means of forecasting expenses.

Background

Cost estimating is a critical function in Air Force weapon system acquisitions. Highlighting its prominence, DoD mandates all programs receive certification of affordability to Congress in order to proceed through the Defense Acquisition System (DAS) Milestone process. The certification of affordability is given by the Milestone Decision Authority (MDA) with concurrence from the Director of Cost Assessment and Program Evaluation (DCAPE) (DoDI 5000.2,2010). The certification of affordability is

granted only after at least two cost proposals are submitted to the CAPE from the System Program Office (SPO) and the Air Force Cost Analysis Agency (AFCAA) (DoDI 5000.2,2010). The CAPE then determines the most realistic cost of the program and makes an affordability decision. Despite this vast rigor and oversight, the Air Force continually underestimates acquisition programs' cost (Arena and others, 6:2006). Critics arguments include the topics: congressional rent seeking; program managers lobbying for jobs; government contractors superior business skills; abundances of federal regulations; unrealistic expectations; underdeveloped technologies; and overly optimistic assumptions (McNaughter, 1989:2) (Cowen and Lee, 1992, 219) (Lee, 1990: 129). The ability to accurately measure expenses that should occur in the future is a challenging concept. In order to capture the ambiguity in predicting the future, cost estimators apply different methods to incorporate reality including: standard deviation, variance, Monte Carlo simulation, and coefficient of variation. While each of these measures has a purpose, this research focuses on the coefficient of variation. The coefficient of variation (CV) is a measurement of dispersion around the mean. It is defined by the standard deviation divided by the mean (Hald, 1952: 77).

$$CV = \frac{\sigma}{\mu} \tag{1.1}$$

where

CV = Coefficient of Variation

μ = mean

σ = standard deviation

It influences cost estimators because they attempt to capture a realistic estimate with a range of uncertainty. An accurate assessment of the uncertainty will fall in a particular CV range. If the estimate's CV is higher than the CV range, it informs decision makers that there is significant risk in the program. If the estimate's CV is lower than the CV range, it informs decision makers that the program estimate may be overly optimistic unless there is sound reason.

Problem Statement

The DoD mandated that risk analysis be incorporated as standard cost estimating practice in 1970 (Arena, 2006:2). However, it was not until 2007 that the Air Force provided coefficient of variation (CV) standards to guide cost analysts. These standards were developed by the Air Force Cost Analysis Agency (AFCAA), a field operating agency whose mission is to perform independent cost and risk analysis and provide special studies to aid long-range planning (Air Force Magazine, 2011:63). As part of their charter, AFCAA conducted a risk study on behalf of the Air Force and published the results in the 2007 version of the <u>Air Force Cost Risk and Uncertainty Handbook</u> (AFCRUH). During this time, AFCAA began using CV as an evaluation criterion when reviewing Program Office Estimates (POE). AFCAA questions the validity of the cost estimate if the POE is outside the published ranges. See Table 1.1.

Table 1.1 CV Ranges Published in AFCRUH

Program Type		
Space Systems	Aircraft Systems	Electronic Systems
35-45%	25-35%	10-20%

AFCAA determined the CV ranges on estimates for a given program are 10-20% on large scale electronics systems, 25-35% on aircraft systems, and 35-45% on space systems (AFCAA, 2007:26).

The published ranges are associated with previous cost estimate performances in the respective programs. The ranges are guidelines for the program throughout the lifecycle and remain stagnant and unchanged. The concern of this research is that early in a lifecycle, System Development and Demonstration for example, it is difficult to accurately capture the uncertainty in a cost estimate. The uncertainty associated with the cost of the system should change as time progresses. As a system matures in development and production, more information is gathered which aids cost estimators in mitigating the uncertainty in the estimate. Thus, the standard for CV should theoretically decrease as time progresses. Brian Flynn and Paul Garvey, conducting research for the Naval Center for Cost Analysis, supported this claim with their study on Department of Navy acquisition programs in 2011. Additionally, Flynn and Garvey found that CV ranges were consistent for all systems regardless of program type (Flynn and Garvey, 2011:29). Their research varies from the Air Force study which suggests different CV ranges for each type of program (space, electronics, and aircraft).

The studies conducted by AFCAA and Flynn and Garvey used Selective Acquisition Reports (SAR) as the basis for their data (Flynn and Garvey, 2011:21) (AFCAA, 2007: 26). The problem with this approach is SAR data provides only a point estimate for budgeting purposes of a system. The uncertainty metrics: Monte Carlo

distributions, confidence intervals, standard deviation, and coefficient of variation are not provided in the SAR.

Research Focus

The previous methods used by Flynn and AFCAA to analyze CV utilized data from SARs. Due to the inherent problems associated with SAR data, this analysis will take a different approach by analyzing cost estimates obtained directly from the individual program offices. This method of data collection removes the interpolation of uncertainty metrics, like coefficient of variation, contained in SARs. The data from the program offices includes all of the uncertainty metrics employed by cost estimators. Additionally, the integrity of primary data from program offices provides validity to the analysis.

The use of primary data provides reliability to the analysis, but also introduces a few limitations. The advantage of utilizing Selective Acquisition Reports is that they are centrally located and easily accessible. Using data from the Air Force systems centers (aerospace, electronics, and space and missile) means three different product centers gather and provide data for the study. The data for this analysis is limited to the amount available at the three product centers. As such, all acquisition programs are not used to formulate the conclusions. This research is limited to the ACAT-I programs from each of the program offices. Currently, there are 51 ACAT-I programs in the Air Force (DAMIR, 2012). This research examines 30 of the available programs in the Air Force.

Another limitation of this study is it focuses only on coefficient of variation. There are several other factors used to capture uncertainty in a cost estimate; however,

the growing popularity of cost estimators focusing on CV and the recent emphasis on CV from AFCAA makes this metric important for cost estimators. This study does not analyze any of the other uncertainty metrics due to the size of the study required to accomplish such an analysis.

Research Questions

The following research questions are investigated:

1 – Does analyzing the coefficient of variation ranges from Selective Acquisition Reports and Program Office Estimates match the coefficient of variation ranges provided by AFCAA in the AFCRUH?

2 – Should there be different coefficient of variation ranges for dissimilar platform types (aerospace, electronics, and Space and Missiles) for Air Force programs?

3 – Do coefficient of variations for Air Force programs change over time?

Model and Implications

This study employs paired t-tests and Tukey methods to capture the accuracy of the factors contributing to a significant coefficient of variation range. The analysis includes paired t-tests to measure the change in CV over time. The outputs of the Monte Carlo simulations conducted by the program offices and the CVs calculated from the SARs serve as the inputs to the models. The conclusion as to whether or not CVs decrease over time in Air Force programs is dependent upon the paired t-test results.

The Tukey method is used to uncover whether or not there should be different CV ranges for dissimilar platform types. It is a multiple comparison procedure used to indentify statistical differences among means (Peck and others, 2001:768). The Tukey method analyzes the differences in means for the CVs to determine if the platform types should be categorized differently. The results of the Tukey method determine if the different platform types should have separate CV ranges. Once the Tukey method is applied, the CV ranges for each group are captured. These CV ranges reflect the recommended CV range for each platform type.

The value of cost estimates is measured by the utility of the decision maker. Cost estimates serve as one of many tools available to decision makers who balance resources to accomplish the mission. By increasing the integrity of the uncertainty captured in cost estimates, decision makers will have more faith in the estimates. The decision makers will gain confidence in cost estimators when deciding which programs to fund since the cost estimates will represent a more realistic picture of a program over its entire lifecycle.

Summary

Department of Defense budgets are decreasing. The value of cost estimators is increasing as the Air Force uses taxpayer dollars more diligently. Cost estimators are employed to provide an accurate assessment of future costs of resources to decision makers. One of the measures to capture the uncertainty of a cost estimate is the coefficient of variation. This analysis will utilize paired t-tests and Tukey analysis to find the most constructive range for CV throughout the acquisition lifecycle.

The remainder of this thesis is divided into four additional chapters: literature review, methodology, results, and conclusions. Chapter two, the literature review, examines the value of cost estimating, the role of uncertainty within cost estimating, and the previous research conducted on CV. Chapter three, methodology, explains in detail the techniques used to analyze the CV. The purpose of this chapter is to provide the reader with a step-by-step method in order for them to reiterate the process and achieve the same results. Chapter four explains the findings and significance of the proposed research questions. Lastly, chapter five concludes the research and provides the practical implications for the cost analyst. It also provides the framework for future research concerning this topic.

II. Literature Review

This chapter is an overview of related topics and previous research. This literature review focuses on the relevance of cost estimating, the importance of capturing uncertainty in cost estimating, and the function of coefficient of variation in cost estimates. The following sections provide brief descriptions of the literature that the researcher reviews to conduct the analysis. The topics of the literature review provide the reader with an understanding of the scope of the study.

Cost Estimating

Cost estimating is a discipline focused on collecting and analyzing data using quantitative and qualitative techniques to forecast costs which aid decision makers with allocating resources. It is both an art and science because of the limited information, variety of techniques, and importance of communication that is attributed to the estimate (Air Force Cost Analysis Handbook, 1-3:2007). The value of cost estimating is reflected in the legislation which mandates cost estimates be conducted for Major Defense Acquisition Programs (MDAP). An MDAP is not allowed to proceed to the next phase of the acquisition process without approval of the milestone decision authority whose performance is reported to Congress (DoDD 5000.1, 4:2007).

The milestone decision authority determines the "affordability" of an MDAP at Milestone B and Milestone C, shown in Figure 2.1, based on cost estimates which include total life-cycle or, if available, total ownership cost (DoDI 5000.2, 23: 2008). Total life-cycle costs include the expenses incurred for conceptual analysis, technological development, requirements planning, acquisition, and operations and maintenance (GAO, 2009:1). The life-cycle costs capture all funds incurred for developing, operating, and disposing of a weapon system.

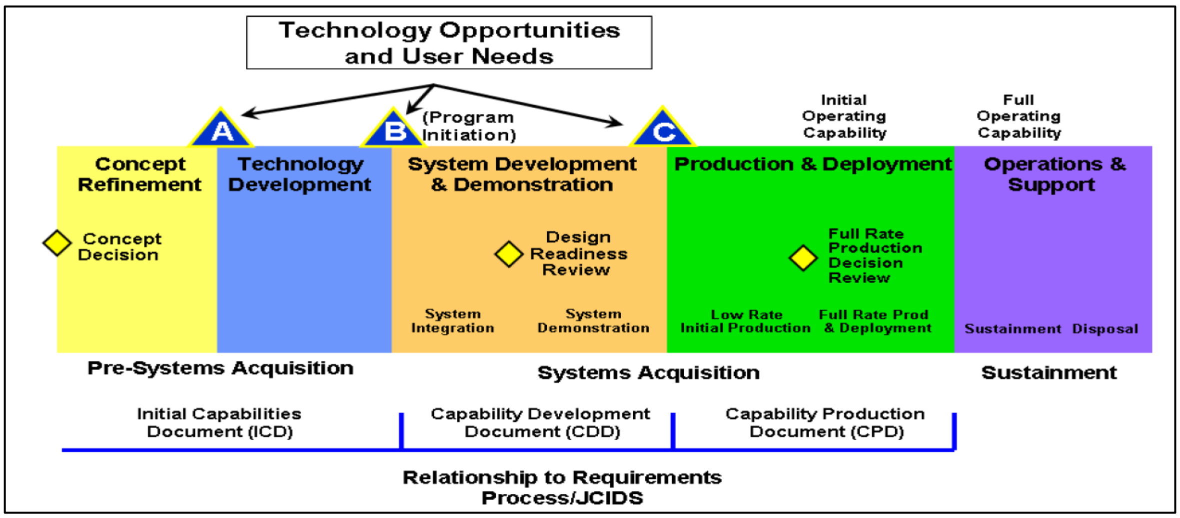


Figure 2.1 Defense Program Acquisition Framework

Affordability

The affordability statement's purpose is to ensure a MDAP fits in DoD long-range plans, and the resources are available to fully fund the program for its entire lifecycle (Defense Acquisition Guidebook 3.2; 2010). All participants in the acquisition process need to consider cost and performance independently to ensure DoD can afford a program beyond the procurement effort. Therefore, the affordability assessment cannot be completed without an estimate of the entire lifecycle (Defense Acquisition Guidebook,

3.2.4: 2010). The affordability assessment is based on a point estimate of the system's costs. Although cost estimators attempt to capture risk and uncertainty, the final decision is based off of one number produced by the cost estimators.

Point Estimates

A point estimate represents a number within a range of possible values representing the total life-cycle cost of a program (AFCAA, 2007:1). The point estimate in the DoD cost community is the best estimate of a system and its requirements minus risk and uncertainty. A point estimate starts with a program manager approving a Cost Analysis Requirements Description (CARD) developed by cost estimators. The cost estimators rely on engineers, program managers, and developers as the technical experts when constructing a CARD. The CARD represents the Work Breakdown Structure (WBS) of a program with associated costs for each element. The arithmetic sum of each program element in a CARD represents the Technical Baseline Estimate (TBE). A TBE is a point estimate; however, it does not typically represent the point estimate chosen as the baseline of a program. It represents the arithmetic sum of most likely values for each WBS element. The TBE is a traceable reference point on which the cost risk analysis is anchored (AFCAA, 2007:2).

The value of each element in the CARD is derived from different estimating techniques including: analogy/factors, parametric, engineering build-up, extrapolation from actual, and Subject Matter Expert (SME) (AFCAH, 2007:Ch3, 1). The analogy/factor method uses costs of similar systems previously developed as a tool to estimate the cost of the weapon system currently being developed. The parametric

method uses Cost Estimating Relationships (CERs) based on historical data to estimate the project's cost. The parametric method applies cost drivers, such as weight and size, to derive the cost of the element in the WBS. The engineering build-up method sums the costs from the lower levels of the WBS to provide a traceable estimate for the WBS element. The extrapolation from actual method uses data already obtained from the current development to estimate future expenses; learning curves are an example of actual data extrapolation. Lastly, the Subject Matter Expert method involves asking professionals closely related to the activity for their input for forecasting costs of the WBS element (AFCAH, 2007:Ch3, 1) (GAO, 2009:107-112).

The risk-adjusted position of a program estimate incorporates cost risk analysis methods which add risk and uncertainty to the point estimate**.** The cost estimators capture risk and uncertainty in the estimate by applying simulation techniques to individual elements in the CARD. Finally, one number is selected as the estimate for a program based on the most realistic assumptions available at the time. The program estimate is selected from a cumulative distribution function derived from Monte Carlo, Crystal Ball®, or Latin Hypercube simulation techniques (NAVSEA, 2005: 4-24). Generally, the mean is selected as the point estimate which is approximately the 50-60 percent confidence level (AFCAH, 2007: Ch 11, 5). However, some program offices, Aerospace Systems Center for example, previously elected to use the 90 percent confidence level to capture more risk and uncertainty when selecting the point estimate (Hudson, 2005). As of February 2013, program offices have elected to evaluate

programs at the mean. The standard is constantly changing. Figure 2.2 shows the output of a Monte Carlo simulation and a selected risk-adjusted estimate.

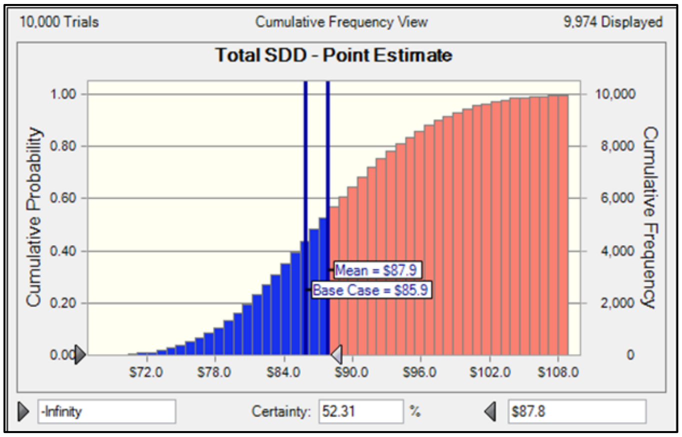


Figure 2.2 Point Estimate

Figure 2.2 depicts a range of possible estimates. A more conservative program manager would choose a value with a higher probability of success. The problem with providing decision makers with only a point estimate is it can be deceiving. The Government Accountability Office found unrealistically low estimates for space acquisition systems and Navy Shipbuilding Programs, in part because of poor choices on the selection of the risk-adjusted estimate (GAO, 2006: 13) (GAO, 2005: 5) .

The derivation of a point estimate is not a clearly defined process. There is no standardized guidance on the selection of a risk-adjusted estimate. The estimate can represent the 'most likely' cost (mode), the 50% confidence cost (median), the 'average' cost (mean), or any other descriptive statistic believed to be the most realistic representation of a program's expected cost. The uncertainty and confusion as to what a point estimate truly represents makes it virtually useless to decision makers (Book, 2004). Figure 2.3 represents different 'most likely' costs of programs with different distributions

attributed to the estimate. Figure 2.3 depicts the uncertainty and confusion as to what point estimates represent.

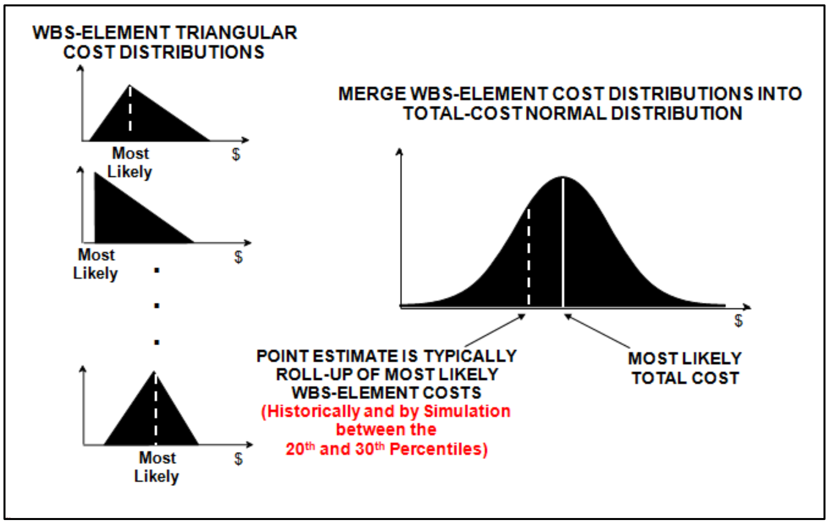


Figure 2.3 Most Likely Cost Estimates (Book, 2004)

The point estimate is derived from costs listed on the CARD developed by cost estimators, engineers, program managers, and developers. The CARD does not contain all necessary information to make a realistic cost estimate though. The CARD does not take into account the risk of building a system (Book, 2004). The CARD is a technical description of the program and is not used to list the associated risks of a program. The risk-management plan should be used to build a cost estimate in conjunction with the CARD. The risk-management plan lists risk issues that could cause problems during development and increase the expected cost. The risk issues are not listed in the CARD because they are not certain; however, if one of the listed risk factors occurs during development it impacts the cost of procuring the weapon system (Book, 2004). A program's cost is not well represented by any singular number. A cost risk analysis

should be conducted and briefed to decision makers to provide more valuable information as to what the risk adjusted estimate accurately represents (Book, 2004).

Point estimates often give decision makers little valuable information about the likelihood of success of an estimate (GAO, 2009: 21). Due to the inherent nature of forecasting, it is challenging to accurately assess the cost of a program years before it is completely developed, manufactured, and disposed. Providing decision makers with only a single value as the estimate is one reason DoD acquisitions struggles with cost and schedule overruns in defense procurement projects.

Cost Growth

Cost growth in defense acquisitions is the difference between the final cost of a program and the estimated cost of a program using Milestone B estimates. It is usually discussed in terms of a metric called the Cost growth Factor (CGF) which is a ratio of the final, or most recent, cost divided by the Milestone B estimate (Arena and others, 2006 :19). A CGF less than 1.0 indicates a program that cost less than initially budgeted. A CGF greater than 1.0 represents a program that has overrun its budget.

Cost growth has been studied for decades by various institutions; however, the three primary contributors to cost growth studies are RAND Corporation, Institute for Defense Analyses, and U.S. Government Accountability Office. The cost growth studies have historically used Selective Acquisition Report (SAR) data to evaluate the cost growth on defense programs. The studies use the published SAR data to calculate the CGFs of different programs.

As early as 1950 researchers found inaccurate cost estimates (Alchain, 1: 1950). In 1950, RAND measured the reliability of different cost estimates. RAND measured the accuracy of engineering estimates, cost estimator's reliability, public engineer's construction cost estimates, and the variations among contractor's bids. Figure 2.4 summarizes the variance of the cost estimate accuracy in RAND's 1950 study.

Study 1		
	Wing Weight Estimates Based on Design	± 10
Study 2		
	Cost estimator's estimates of cost -- one producer	± 23
Study 3		
	Public Engineer's cost estimates of Construction Projects	± 15
Study 4		
	Variation Among Contractors' Bids	± 21

Figure 2.4 1950 RAND Cost Reliability Results

The purpose of the study was to quantify the reliability of different types of cost estimates. RAND showed with 90% confidence in the 1950s that initial cost estimates vary between 10 and 23% of the actual cost. RAND identified unclear specifications, changes to specifications, and variance among manufacturers as the primary reason for the estimating differences. The study shed light on a topic that continues to be studied 60 years later.

GAO Space Acquisition Cost Growth

A U.S. Government of Accountability office publication found original cost estimates for space programs increase by 44 percent (GAO, 2006: 1). The study focused

on six major space acquisition programs in the Air Force. The GAO used a case study methodology to examine which areas in cost estimates for space system acquisition have been unrealistic and what incentives and pressures contributed to the quality and usefulness of cost estimating. The results of the analysis showed a tendency for Air Force to start space acquisition programs with unrealistic requirements because of pressures to secure funding. The study found the program office estimates were too optimistic and the Air Force did not rely heavily enough on the required Independent Cost Estimates (ICE) (GAO, 2006: 32-37). It appears the program office estimates were selected as the baseline estimate because they were lower than the ICE and more likely to acquire funding. Figure 2.5 shows three baseline estimates where the program office estimate was lower than the ICE due to unrealistic assumptions in order to secure funding (GAO, 2006: 37-40).

Table 7: Comparison of 2004 AEHF Program Office and Independent Cost Estimates

Millions of fiscal year 2006 dollars

Program office estimate	Independent cost estimate: AFCAA	Independent cost estimate: CAIG	Difference	Latest program office estimate
$6,015	*	$8,688	44%	$6,132

Source: CAIG and GAO analysis.

Table 9: Comparison of 2003 NPOESS Program Office and Independent Cost Estimates

Millions of fiscal year 2006 dollars

Program office estimate	Independent cost estimate: AFCAA	Independent cost estimate: CAIG	Delta	Latest program office estimate
$7,219	$8,869	*	23%	$11,400

Source: Air Force Cost Analysis Improvement Group briefing, April 2003.

Table 11: SBIRS High GEO 3–5 Procurement Funding Analysis

Millions of then-year dollars

Cost baseline	CAIG estimate	Program office	Delta	Delta %
Three individual satellite procurements	$2,892	$2,027	$865	43%

Source: CAIG and GAO analysis.

Figure 2.5 Optimistic Cost Estimates for Space Systems

IDA Major Causes of Cost Growth

A study conducted by the Institute for Defense Analysis (IDA) cited poor management and weak program definition (Porter and others, 23: 2010). The IDA was sponsored by the Office of the Director Acquisition Resources & Analysis (OUSD AT&L) and tasked to seek a deeper understanding of the decisions and mistakes that contribute to cost growth. The IDA studied 11 programs that entered full-scale development and experienced significant cost growth. The study team reviewed cost history data and interviewed cost estimators and senior acquisition officials. The study relied heavily on SAR data, typical of the majority of cost growth studies conducted in the defense industry. The IDA claimed poor acquisition management was responsible for inappropriate implementation of policies. The weak program definition led to unstable requirements, decisions based on immature technologies, and excessive schedule compression (Porter and others 23, 2010). The findings of the IDA study are shown in Figure 2.6.

1. Weaknesses in management visibility, direction, and oversight

a. Lax or inappropriate implementation of policies

b. Excessive reliance on unproven management theories and acquisition strategies

c. Poor contractor selection, oversight, and incentivization

2. Weaknesses in initial program definition and costing

a. Defective and unstable requirements processes

b. Entry into development with immature technologies

c. Deficient front-end analysis of system-level design issues and technical risks

d. Excessive schedule compression and concurrency

Figure 2.6 Major Causes for Cost Growth (Porter and others, 2010: 24)

Sources of Cost Growth

More research cited poor cost estimating and increases in requirements lead to cost growth during the development phase of the acquisition lifecycle. Quantity changes are responsible for procurement cost growth; while the largest contributor to cost growth is poor managerial decisions (Bolten and others, 46:2008).

No matter the cause, all poor estimates lead to unrealistic budgeting and underfunded programs (McNicol, 9:2004). The studies suggest poor initial cost estimating leads to cost growth. Even though the topic is analyzed extensively, the trend of cost growth remains high with little significance of improvement (Younossi and others, 45: 2007).

Cost Growth Trending

Cost growth is not a new problem for DoD. Cost growth in defense acquisitions has been studied for over fifty years (Fox and others 2001:7). As far back as the 1950s, the President, Congress, Secretary of Defense, and service chiefs have launched initiatives to curb cost growth through acquisition reforms. In the 1950s and 60s business executives Robert McNamara and David Packard launched management initiatives to centrally control acquisition decisions (Fox and others, 2011:43). McNamara and Packard's influence on the defense acquisition process can still be seen in the current structure.

In the 1970s the Blue Ribbon Defense Panel was established to identify the reasons for defense acquisition cost growth and schedule overruns. The Blue Ribbon Defense Panel stood up several government agencies and policies to improve the defense

acquisition process which are still functioning today: Defense Systems Acquisition Review Council (DSARC), Cost Analysis Improvement Group (CAIG), DoD Directive 5000.1, DoD Instruction 5000.2, and Defense Federal Acquisition Regulation (DFAR) 2011, (Fox and others, 2011:62-95).

The 1980s was a period that experienced substantial defense budget increases with the help of President Ronald Reagan. The Regan administration increased defense procurement budgets by as much as sixty-one percent (Fox and others, 2011: 101). The increased budgets were coupled with fewer restrictions which are believed to be a direct contribution to many allegations against the DoD for fraud, waste, and abuse. Much of the acquisition reforms of the 1980s were initiated to curtail the fraud, waste, and abuse allegations. The 1980s is responsible for producing the Title V of the Goldwater-Nichols Act of 1987 which prompted a division of labor between acquisition management and support functions at the command level (Fox and others, 2001: 138). Also, the Nunn-McCurdy Amendment was put in place which requires a notification to congress if there is cost growth greater than fifteen percent and a termination of the program if cost grows by more than twenty-five percent, unless the secretary of defense can provide a detailed explanation certifying the program is essential (Fox and others, 2011: 120).

The 1990s were focused on introducing a more responsive, effective, and efficient approach to defense acquisition. There were more than sixty-three acquisition reforms in the 90s (Hanks and others, 2005:94). Several of the key acquisition reforms include: Defense Acquisition Workforce Improvement Act of 1990 (DAWIA), Federal Acquisition Streamlining Act of 1994 (FASA), Federal Acquisition Reform Act of 1996

(FARA), Cost as an Independent Variable (CAIV), and the Clinger-Cohen Act of 1996 (Hanks and others, 2005: 1994).

The 2000s remained relatively quiet until 2009 when the Weapon System Acquisition Reform Act (WSARA) was signed into law by President Barack Obama. Before WSARA, major program spending limits, phases and milestones were redefined. The Joint Capabilities Integration and Development (JCID) process was introduced along with the Defense Acquisition Guidebook (DAG) to continue a history of introducing acquisition legislation (Fox and others, 2011, 225-227). WSARA was the largest piece of acquisition legislation introduced in the twenty-first century. WSARA aims on increasing focus on trading off cost, performance, and schedule, increasing systems engineering efforts earlier in the program lifecycle, providing clearer guidance on analysis of alternatives and cost estimating procedures, increasing competition throughout the program lifecycle, and restricting the organization, including the appointment of several new administrative officials (DAU, 2010).

Despite the dozens of acquisition reforms and legislation changes, the explanations for cost growth remain focused around the same five principles first identified in the 1950s: 1) schedule slippage, 2) lack of qualified personnel, 3) high turnover frequency, 4) inadequate cost estimating, 5) insufficient management on contractor performance, 6) unclear requirements definition (Fox and others, 2011, 35). Cost growth is and has always been a problem. Realistic cost estimates will allow decision makers to make more informed decisions when choosing among major weapon systems. In order to educate decision makers, it is important to provide them with a cost

range and not a single estimate. A cost estimate is not an absolute number which will remain constant (Fisher, 1:1962). An estimate is based on agreed upon assumptions that cannot all be true, but rather, as accurate as possible. It is imperative that risk and uncertainty are captured in all cost estimates.

Risk and Uncertainty

Although the terms risk and uncertainty are commonly interchangeable in casual conversation, the two concepts are unique for this research. Risk is defined as the chance of loss or injury. Uncertainty is the indefiniteness about the outcome of an event (Air Force Cost Risk and Uncertainty Handbook, 4: 2007). It is extremely unlikely that the forecasted number will actually reflect the true cost of a weapon system. The lack of knowledge about the future is only one reason for the difference. Equally important reasons are inaccuracies in historical data, poor assumptions, equations, and relative factors used to derive an estimate (GAO Cost Estimating and Assessment Guide, 153: 2009). The inabilities of cost estimators to accurately estimate MDAPs are evidence of the need to capture uncertainty around a point estimate.

Less information is known about a program early in the lifecycle. As a MDAP progresses through the acquisition lifecycle, more data is collected that accurately reflects the outcome of the program. Therefore, cost estimates are more accurate later in the programs lifecycle (Arena and others, 39: 2008).

The changes in uncertainty of a cost estimate are reflected in Figure 2.7 (GAO, 2009:155).

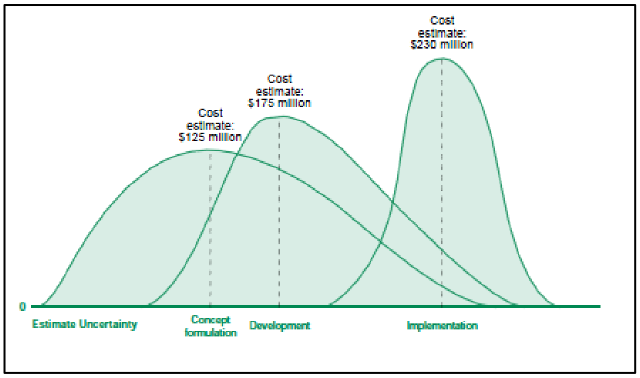


Figure 2.7 Changes in Cost Estimate Uncertainty across the Acquisition Lifecycle

It is important to communicate to decision makers that there is more uncertainty in the point estimate earlier in the MDAP lifecycle.   Cost estimators present their data as point estimates, but also include various descriptive statistics to capture risk and uncertainty to communicate the likelihood of the program overrunning.

Descriptive Statistics

Descriptive statistics are valuable for portraying the risk and uncertainty in cost estimates. Cost estimators often choose to brief decision makers with different statistics to represent uncertainty including: mean, median, mode, confidence interval, standard deviation, cumulative distribution functions, correlation, and coefficient of variation (AFCAH, 2007: 94). A cost estimator uses the different statistics to portray various characteristics of the estimate. To capture risk and uncertainty, an estimator produces multiple estimates or simulates different 'what-if' scenarios (GAO, 2009: 185). The mean is the average of all estimates divided by the number of trials. The mode is the most

common estimate of all trials. The median is the middle value of all trials. In Monte Carlo simulation the estimate is usually replicated between 10,000 to 100,000 times, which is simple to produce with modern computing capabilities. As the number of trials increases in size the mean, median, and mode converge until:

$$\text{Mean} = \text{Median} = \text{Mode} = \mu = \frac{1}{n}\sum_{i=1}^{n}(x_i - 1)^2 \qquad (1.2)$$

This represents the 'most likely' value which represents the 50 percent confidence level (GAO, 2009:167).

The confidence level is the percent of certainty in the estimate. It represents an interval around the mean (Sachs, 1982: 112). An 80 percent confidence level represents a value where 80 percent of the Monte Carlo simulations produced an estimate at that value or lower. To a decision maker, an 80 percent confidence level depicts an estimate that has a 20 percent chance of exceeding the budget (AFCAH, 2008: 11-13). Figure 2.8 shows a cumulative distribution function (CDF) and previously mentioned statistical parameters.

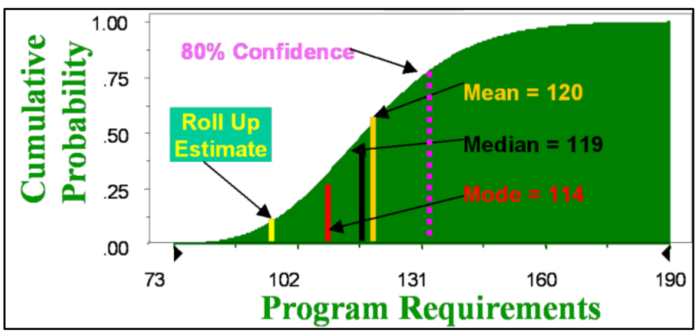


Figure 2.8 Cumulative Distribution Function

A cumulative distribution function is commonly referred to as an S-curve (AFCAH, 2008: 11-13). A Monte Carlo simulation produces a CDF depicting the different parameters (Dienemann, 12: 1966). The CDF represents the probability that a random variable assumes a value less than or equal to the given confidence level (Sachs, 1982: 44).

Standard deviation is another statistical parameter commonly used by estimators to scale risk and uncertainty. The standard deviation is used to determine the amount of dispersion around the mean of a given data set (GAO, 2009: 97). Larger standard deviations in estimates represent larger uncertainty. The standard deviation essentially measures the average distance between data points and the mean (AFCAH, 2007: 64). It is valuable for analyzing data points in the same data set; however, to compare variances between different data sets coefficient of variation is a more effective measure.

Coefficient of Variation

The coefficient of variation (CV) is becoming one of the most recognized metrics to characterize cost estimating risk and uncertainty distributions (AFCAH, 2007: 64). It is defined by the standard deviation divided by the mean (Sachs, 1982: 77). The CV is useful for comparing variances between data sets. In essence, CV normalizes the risk and uncertainty in estimates among various programs (GAO, 2009: 98). The CV is useful for comparing variability among program types. It may be known from historical estimates that aerospace programs typically have an uncertainty range represented by 30 percent variability in cost. If an estimator produces a point estimate with a CV lower than 30 percent, it flags decision makers that there may be overoptimistic assumptions in the

estimate, or at least some justification should be provided for the abnormally low CV. Categorizing appropriate CV ranges for different programs at particular points of the acquisition lifecycle is important because they are easy for decision makers to comprehend.

Previous Research

The Air Force Cost Analysis Agency categorized CV ranges for Air Force programs in a study conducted in conjunction with Tecolote Research, Inc. for the 2007 version of the Air Force Cost Risk Uncertainty Handbook. The CV ranges were derived from a study of Selected Acquisition Report data on completed Air Force programs. The details of the methodology are not disclosed, but AFCAA acknowledges that the results are consistent with observed rules-of-thumb. AFCAA concedes further study is needed to produce higher fidelity in their recommendations (AFCAA, 2007: 26).

In 2011 the Naval Center for Cost Analysis (NCCA) produced a study which analyzed coefficient of variation to determine five conjectures: 1) CVs in current cost estimates are consistent with those computed from acquisition histories 2) CVs decrease throughout the acquisition life cycle 3) CVs are equivalent for aircraft, ships, and other platform types 4) CVs decrease when adjusted for changes in quantity and inflation 5) CVs are steady over the long run. The study analyzed 100 naval acquisition programs from Selective Acquisition Reports. The researchers used the baseline estimate and the current estimate to calculate a cost growth factor for each program. The current estimate divided by the baseline estimate is equivalent to the cost growth factor. A distribution was fit around the data points and the coefficient of variation was calculated. The

researchers grouped the data points into categories for time and program type. They wanted to determine if the CVs decrease over time or are similar among program type. The analysis yielded: 1) CVs are historically pervasively underestimated 2) CVs decrease throughout the acquisition lifecycle 3) CVs are equivalent among program type 4) CVs decrease for changes in quantity and inflation 5) CVs are not steady over time (Garvey and Flynn, 2011: 20-29).

The previous research differs because AFCAA determined the CV ranges among program types differ, whereas NCCA found that CVs are equivalent regardless of program type. AFCAA also recommend one range per program type regardless of where the program was in its acquisition lifecycle. NCCA found that CVs decrease overtime and the CV should be adjusted to accurately capture uncertainty.

AFCAA and NCCA studies utilized SAR data to conduct the analysis, as have most cost growth studies. The problem is that SAR data are usually inaccurate. The estimate in SARs does not always equal the Program Office Estimate (POE), the Independent Cost Estimate (ICE), or Non-Advocate Cost Assessment (NACA). These estimates are typically the final estimate derived from sources most familiar with the program. The current estimate in a SAR aligns with the President’s Budget submission. The budget submission is the amount programmed for the MDAP but it does not always reflect the forecast of the cost estimators.

A study found that SAR data fail to use consistent baseline costs, exclude significant elements of cost, exclude classes of major programs, change preparation guidelines, inconsistently interpret preparation guidelines, produce unknown and variable

funding levels for program risk, share costs in joint programs, and report the effects of cost changes rather than their root causes (Hough, 1992: 5). These inaccuracies reflect poorly on the quality of the SAR database. The imprecise data found in SARs does not invalidate previous cost growth studies; it merely reinforces the need for caution when examining the results of the studies (Hough, 1992: 42). The best source of data for individual weapon system remains with the program offices (Hough, 1992: 42)

Conclusion

This chapter provides an overview of related topics and previous research. The literature review begins with an overview about the importance of cost estimating in DoD acquisitions. The importance of capturing risk when briefing decision makers about estimates is then discussed. Finally, we reviewed the previous literature concerning descriptive statistics, specifically coefficient of variation. The goal of this chapter is to provide the reader with the scope of this study. The next chapter, the methodology, presents the step-by-step directions to reenact the analysis of the researcher. The limitations and assumptions of this study are discussed in detail. Subsequently, the model used to determine the optimum range for coefficient of variation throughout the acquisition lifecycle is presented.

III. Methodology

This chapter describes the data used for determining the optimal range for the coefficient of variation in cost estimates at different stages in a program's acquisition lifecycle. The limitations and assumptions are described in detail. Last, the theoretical and practical application of the processes and procedures for conducting the analysis are detailed, which provides the reader the ability to replicate the analysis.

Data Source

The primary data for this analysis come from four acquisition product centers around the United States Air Force, and the secondary data come from the Defense Acquisition Management Information Retrieval (DAMIR) database. Aerospace Systems Center, Air Armament Center, Space and Missile Center, and Electronics Systems Center provided the primary data for the analysis. The secondary data come from DAMIR which is a DoD initiative that provides enterprise visibility to acquisition program information. DAMIR is managed and operated by the Office of the Secretary of Defense for Acquisition, Technology and Logistics/Acquisition Resource Analysis.

Primary Data

The primary data are PowerPoint® briefing slides that are developed for the program office estimate (POE) or the independent cost estimate (ICE). The slides are used to brief either the Air Force Cost Analysis Agency (AFCAA) or the Air Force Cost Analysis Improvement Group (AFCAIG) during the annual program reviews. The slides include the current status of the program, the current point estimate and risk range, and

future outlook of the program. An example of the slides is attached in Appendix A. The slides are unique to this analysis because they contain the risk and uncertainty ranges of the cost estimate each year. The annual replication of the slides provides an update to the changes in the uncertainty of the program and insight to the overall progress. Also, the briefings are derived by the program office cost estimator and program manager who possess first-hand knowledge of the program.

The Powerpoint® slides were reviewed for specific information and not all presentations contained the same categories of information. The categories of information along with a description of the categories are shown in Table 3.1. The most critical piece of information needed for this analysis was the CV calculated by the program office. Appendix A shows some examples of the specific information utilized for the analysis.

Table 3.1 Powerpoint® Slide Information

Category	Description
Year	The year the presentation was developed
Platform Type	The type of weapon system. Ex. Avionics, Engine, Plane, Satellite
CV	The coefficient of variation of the risk analysis
Mean	The mean estimate oof the risk analysis
Standard Deviation	The standard deviation of the risk analysis
Lifecycle Location	The specific location of the program. Ex. Milestone B+3 is the program 3 years past MS B
Milestone Location	The current location of the program. Ex. Milestone B or C
80 % Confidence	The 80% confidence level of the Monte Carlo simulation used in the risk analysis
Program Office Estimate	The point estimate provided by the program office or AFCAA
Base Year	The year the base line estimate was established. Usually the date of Milestone B
Estimate Dollars	Base Year or Future year for the program office estimate
Lifecycle Stage	System Development & Demonstration or Production & Deployment
Program Type	MDAP or MAIS

Typical cost growth studies use Selective Acquisition Reports (SAR) as the primary data source. The SARs do not contain risk and uncertainty ranges. The SARs

typically present a point estimate that has been adjusted by several agencies more distant to the program. The SARs are usually updated after the POE and ICEs are reviewed by the Milestone Decision Authority (MDA). The SARs serve as the source for the secondary data for this study.

Secondary Data

The secondary data are retrieved from DAMIR. DAMIR is an online database that contains DoD acquisition program information. Specifically, the SARs for all Major Defense Acquisition Programs and Major Automated Information Systems are contained in DAMIR. The previous cost growth studies, mentioned in chapter two, used SAR data to perform their analysis. The research of Garvey and Flynn, on the coefficient of variation in naval programs, used SAR data to compute Cost Growth Factors to analyze appropriate CV ranges for naval programs (Flynn and Garvey, 2011:21). Because this research is attempting to replicate Flynn and Garvey's study with Air Force programs, this research will use the SAR data from the programs contained in the primary data. The program offices provide 30 program's slides. The secondary data are the SARs for the same 30 programs. This method provides secondary study to the analysis and compares the results of this study to that of Flynn and Garvey's.

A list of the 30 programs is shown in Table 3.2.

Table 3.2 Data Provided From Program Offices

Program	Product Center	Platform Type	Program Number
JASSM-ER	AAC	Missile	1
B-2 EHF Inc 1	ASC	Avionics	2
C-5 RERP	ASC	Engine	3
C-27J	ASC	Plane	4
C-130 AMP	ASC	Avionics	5
C-130J	ASC	Plane	6
CRH (H-47)	ASC	Helicopter	7
CRH (H-71)	ASC	Helicopter	8
CVLSP	ASC	Helicopter	9
Global Hawk	ASC	UAV	10
HCMC 130J	ASC	Plane	11
LAIRCM NexGen MWS	ASC	Electronic	12
MQ-9 Reaper	ASC	UAV	13
B-2 DMS	ASC	Avionics	14
B-2 EHF Inc 2	ASC	Avionics	15
B-2 EHF Inc 3	ASC	Avionics	16
MQ-1C Gray Eagle	ASC	UAV	17
3 Dim Lng Rng Radar	ESC	Electronic	18
AF-IPPS	ESC	Computer Sys	19
AFNet Inc 1	ESC	Computer Sys	20
AOC Inc 10.2	ESC	Computer Sys	21
ITS Inc 2	ESC	Computer Sys	22
MPS Inc III	ESC	Computer Sys	23
MPS Inc IV	ESC	Computer Sys	24
GPS III	SMC	Satellite	25
SBIRS GEO 1-2	SMC	Satellite	26
SBIRS SFP GEO 3	SMC	Satellite	27
SBIRS SFP GEO 4	SMC	Satellite	28
SBIRS SAR	SMC	Satellite	29
SBSS Block 10	SMC	Satellite	30

The SAR data used in this analysis were retrieved from the Defense Acquisition Management Information Retrieval (DAMIR) database. DAMIR was established to provide top level oversight to congress to report cost updates on all Major Defense Acquisition Programs (MDAP) and Major Automated Information Systems (MAIS) for all of DoD. SARS are supposed to be published every year after a program enters

Milestone B until the program reaches 90 percent completion; however, there are always exceptions. SARs are sometimes not published during election years due to political influences. For example, very few programs published SARs in 2008. Also, some programs elect not to publish SARs if they are about to enter Milestone C or rebaseline to eliminate redundancy, because programs are required to publish a SAR for every new milestone or rebaseline.

This analysis focused on the Cost and Funding section of the Selective Acquisition Reports. An example is shown in Appendix B. The current estimate and the baseline estimate in base year dollars were extracted to calculate the Cost Growth Factor (CGF) as shown in Equation 3.1.

$$\text{CGF} = \text{current estimate/baseline estimate} \tag{3.1}$$

Other sections of the SAR were utilized to gain more knowledge about the status of the program including: Executive Summary, Responsible Office, Threshold Breaches, and Schedule, but the Cost and Funding section was the primary focus area. The researcher was able to gain a greater sense of awareness about the program by combining the information from the SAR with the program office estimate slides.

Data Limitations

The primary data are provided by four separate product centers. All four offices analyze and present their results differently. The methods used to capture uncertainty vary among program offices. The external influence on the cost estimate changes between programs. These external factors lead to the data not being standardized among

program offices. Also, the primary data are peer-review briefings and AFCAIG briefings. The peer review briefs are analyzed at the program office prior to the program office explaining and defending their estimate to AFCAA. After the AFCAA review, the AFCAIG brief is developed and given to the Office of the Secretary of Defense Cost Assessment and Program Evaluation (OSD CAPE) prior to an adjustment to the SAR.

The two sources, peer-review and AFCAIG briefings, employed for the primary data introduce potential error to the analysis; however, the data are more realistic than SAR data because they are produced by sources closer to the program and contain confidence levels and risk analysis. The reason both peer review and AFCAIG briefs are used is there is no standardized data repository similar to DAMIR for cost estimates. Instead, a search through the program office's file archives provides as many briefings as possible to ensure normality in the analysis.

This method of collecting the data is considered a sample of convenience and introduces sources of error. The sample may not be the most accurate representation of the population. Ideally, it is best to test the entire population or, if possible, take a random sample of the population. The limited collection of data at the program offices and AFCAA combined with resource constraints on this analysis make the convenience sample the only feasible alternative. The use of the SARs as secondary data mitigates some of the error added in the analysis and adds validity to the study.

The sampling technique limited the amount and type of data available. The sample yielded 30 programs. The sample breakdown of the programs by platform type is shown in Figure 3.1.

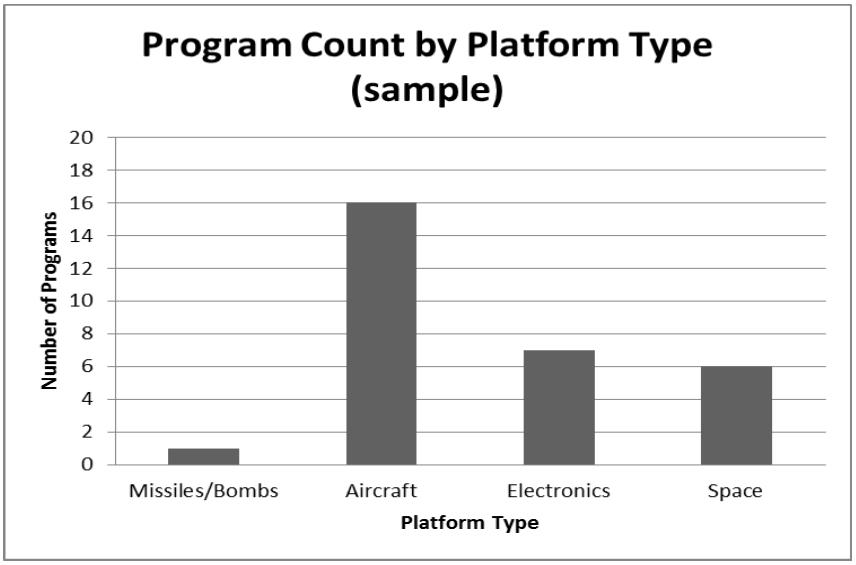


Figure 3.1 Number of Programs by Location in Sample

Figure 3.2 shows the current population of all MDAPs and MAISs by platform type in 2012.

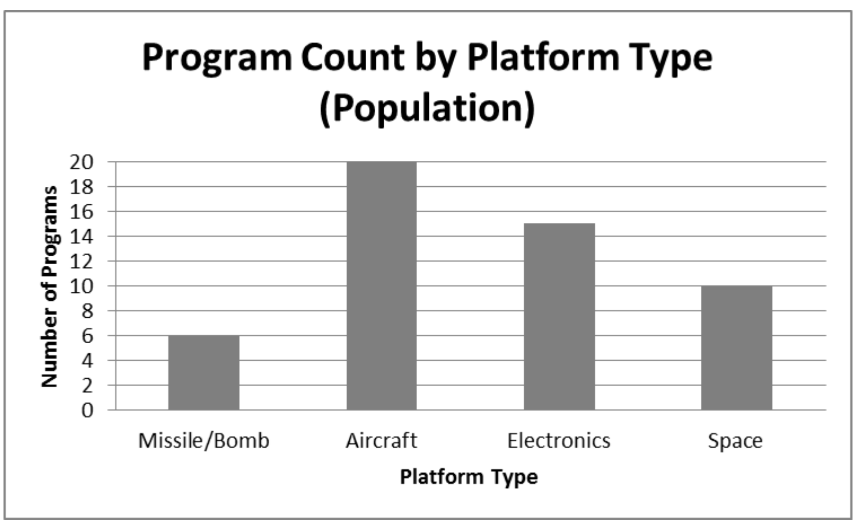


Figure 3.2 Number of Programs by Platform Type in Population

The sample was broken down further to capture programs by platform type. This is done to analyze whether or not the recommended CV range should differ by platform type. Figure 3.3 shows the programs broken down by platform type in the sample.

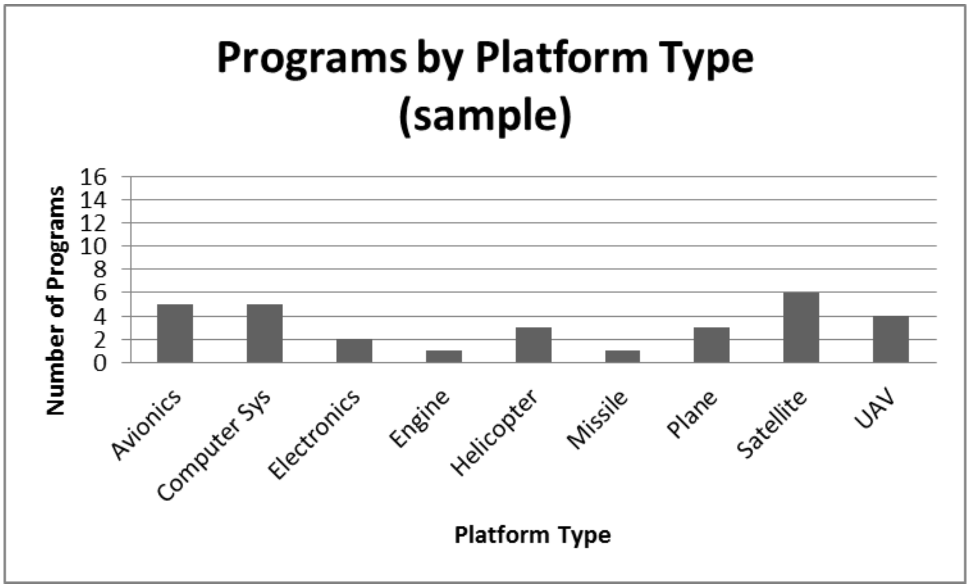


Figure 3.3 Programs by Platform Type in Sample

Figure 3.4 shows the programs in the population broken down by platform type.

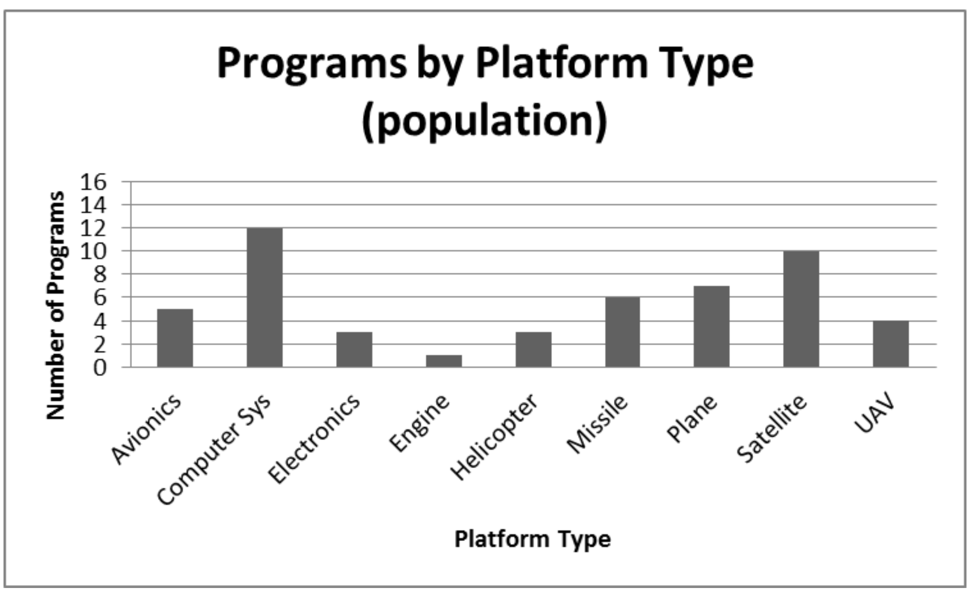


Figure 3.4 Programs by Platform Type in Population

There are two additional programs included in the sample that are not currently in the MDAP or MAIS population. The C-130 Avionics Modernization Program (AMP) was cancelled in the FY13 budget, and the C-27J is expected to be cancelled in the same year and no longer shows up on the MDAP list. Both programs fall under the ASC program location. The C-130 AMP program falls into the avionics platform and the C-27J is included in the plane platform type. These programs were included in the sample because the program office provided historical program office estimates which included the coefficient of variation calculations.

The coefficient of variation calculations were not conducted regularly by program offices until 2007 when AFCAA published the recommended CV ranges for program offices (AFCAA, 2007:26). Therefore, the data were filtered to include only programs that were current as of 2007. This limits the size of the population and reduces the power of the analysis. The sample included as many programs as possible across a range of platform types to reduce some of the error.

Theoretical Procedures and Processes

The goal of this analysis is to answer the research questions developed in Chapter 1. Simplified, the intentions are to recommend CV ranges for Air Force acquisition programs, determine if different CV ranges should be used based on platform type, and determine if CV decreases over the course of the program's acquisition lifecycle.

Chebyshev's Rule

The method used to determine the recommended ranges of CVs for programs is Chebyshev's rule also known as Chebyshev's inequality. Chebyshev's rule is used to

determine the range of CVs because it applies regardless of the distribution of the data. The rule guarantees that in any probability distribution, no more than $1/k^2$ of the distribution's values can be more than k standard deviations from the mean.

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad \text{with } k > 0 \tag{3.2}$$

Therefore, the probability that the absolute difference between a variable and its mean is greater than three standard deviations is no more than $1/3^2$ or 0.11 (Sachs, 1984:64).

$$P(|X - \mu| \geq 3\sigma) \leq \frac{1}{9} = 0.11 \tag{3.3}$$

By using Chebyshev's inequality, the mean and standard deviation of any grouping of calculated coefficient of variations or cost growth factors can be used to calculate a range of CVs that will capture at least 89% of acquisition programs. The range can then be recommended to cost analysts to use when producing cost estimates. Analysts will have confidence that enough risk and uncertainty are included in estimates which is suitable for at least 89% of Air Force programs. The ranges will be operationalized in order to recommend CV benchmarks depending on the type of weapon system or the phase of the acquisition lifecycle depending on the results of this analysis.

Tukey's HSD Test

The method for determining if different CV ranges should be used based on platform type is analyzing the CVs and CGFs through a Tukey's Honestly Significant Difference (HSD) test. The Tukey method is a multiple comparison statistical test. Its purpose is to find means between groups that are statistically significant from each other (Sachs, 1984:534).

The null hypothesis of the Tukey test is that all the means are equal.

$$\begin{aligned} H_0 &= \mu_i = \mu_j \\ H_1 &= \mu_i \neq \mu_j \end{aligned} \tag{3.4}$$

The test statistic for comparing each group to each other is computed by:

$$D = \frac{\mu_i - \mu_j}{\sqrt{MSE / n}} \tag{3.5}$$

where

μ_i = mean of first group

μ_j = mean of second group

MSE = Mean Squared Error

N =number in each group

The test statistic for each group comparison is used in conjunction with the studentized range distribution to test the probability, 1-α, that all differences $\mu_i - \mu_j$ will satisfy the hypothesized inequalities. The degrees of freedom is equal to the total number of observations minus the number of means (Sachs, 1984:534) (Larsen and Marx, 2001:647).

In order for the Tukey HSD test to be valid, three test assumptions must be met. The observations being tested must be independent, normally distributed, and homoscedastic. Independence means the tested variables, CV and CGF, are unrelated in a probabilistic sense. In other words, the occurrence of a previous variable does not affect the probability of the next variable (Sachs, 1984:204). The CV and CGF data will

be tested separately and a comparative analysis will be performed, post hoc, to validate trends in the data.

The normally distributed assumption is important because a Tukey analysis is essentially separate t-tests, discussed later, between the different tested groups. The normally distributed assumption is met if the tested variables are derived from a normally distributed population (Sachs, 1984:58-60). This analysis uses the Shapiro-Wilk goodness-of-fit test to confirm the normality assumption. The null hypothesis of the Shapiro-Wilk test is that the sample comes from a normally distributed population. Therefore, to prove the normality assumption the test must fail to reject the null hypothesis by having a p-value greater than the alpha level of 0.05 (Everitt, 2002:343-344). The Shapiro-Wilk test statistic is shown in Equation 3.6.

$$W = \frac{\sum_{i=1}^{n} a_i {y_i}^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{3.6}$$

where

y_i =the sample data

a_i = the constants to be evaluated

The last assumption needed for the Tukey test to be valid is homoscedasticity or equality of variance. The homoscedasticity assumption is valid if the variables within the group have the same variance. This allows the means between the different groups to be compared for significance (Sachs, 1984:494).

In this analysis, the groups are divided into platform type which is defined by the physical location of the program office. The program offices are located at four separate

product centers: Aerospace Systems Center, Air Armament Center, Electronics System Center, and Space and Missile Center. The product centers represent the type of platform: Aerospace Systems Center represents aircraft programs, Air Armament Center represents missile and bomb programs, Electronics Systems Center represents electronic programs and the Space and Missile Center represents space programs. The groups are defined by the physical location of the program office because the sample is not large enough to separate the programs by the type of weapon system: airplane, helicopter, UAV, electronic, missile, or satellite. Ideally, analyzing the data by the type of weapon system regardless of program office location would be best; however, the sample size in this study limits the capability to analyze the data in this manner. The purpose of separating the data into groups is to see if there are statistically significant differences between the means of the groups. If there are differences in means, then it can be stated that there should be different CV ranges based on platform type. The ranges are determined by the application of Chebyshev's inequality mentioned previously.

Paired t-test

The method used to determine if the coefficient of variation decreases over a program's lifecycle is a paired t-test. The difference from the last CV calculated and the first CV calculated are used as the observations. The t-test is paired because the individual observations, the last and first observation, are as homogeneous as possible (Sachs, 1984:307). The CVs are calculated from the same program.

The t-test is used as a one sided test. The hypothesis is shown in Equation 3.7.

$$\begin{aligned} H_o &\geq 0 \\ H_a &< 0 \end{aligned} \tag{3.7}$$

The test statistic for the paired t-test is shown in Equation 3.8.

$$\hat{t} = \frac{\bar{d}}{s_{\bar{d}}} = \frac{(\sum d_i)/n}{\sqrt{\dfrac{\sum d_i^2 - (\sum d_i)^2/n}{n(n-1)}}} \tag{3.8}$$

For this test, if the p-value is less than the alpha of 0.05 then the test will be rejected. The rejection provides statistically significant evidence that the CV decreases over time. The t-test assumes the differences calculated for each pair, the last and first calculated CVs, are normally distributed. The normality assumption is validated using the Shapiro-Wilk test as mentioned earlier.

Practical Procedures and Processes

JMP® software from SAS is used to conduct the analysis. JMP® is statistical software which combines robust analytics with dynamic graphics to enable visual discovery (SAS, 2012). The data are input into JMP® with each program entered as its own data point. The program represents one data point regardless of how many years of cost estimates are gathered for that program. For example, the C-5 Reliability Enhancement and Reengining Program (RERP) includes estimates for FY04, FY05, FY07, and FY10; however, the C-5 RERP is entered as one data point represented by a single row in JMP®.

The columns in the analysis represent individual characteristics for each program. Each column represents one year of program data. The different columns represent the

categories and calculations including: Year, Platform Type, CV, Mean, Standard Deviation, Lifecycle Location, Milestone Location, 80% Confidence, Program Office Estimate, Base Year, Estimate Dollars, Lifecycle Stage, Program Type. The arrangement of data facilitates simple calculations to answer the research questions.

The first research question is, "Does analyzing the coefficient of variation ranges from Selective Acquisition Reports and Program Office Estimates match the coefficient of variation ranges provided by AFCAA in the AFCRUH?" To test this question a Tukey analysis is performed. A Tukey analysis tests the total number of means in a sample. It is suitable for testing two or more groups of means to determine if there is a difference (Sachs 1982: 534). This calculation will utilize the column "CV" and "BY CGF". If the Tukey analysis shows there is a difference in means depending on platform type or program location (as currently assumed in AFCRUH), then a distribution for these columns will be analyzed and Chebyshev's rule will be applied to the distributions. A range will be calculated using Chebyshev's rule and compared to the current recommendations in AFCRUH.

The Second research question is, "Should there be different coefficient of variation ranges for dissimilar platform types (aerospace, electronics, and Space and Missiles) for Air Force programs?" The Tukey analysis will be used to compare the means based on platform type and program office location. JMP® makes the calculation effortless by plotting a Y by X graph and comparing means across all pairs. JMP® will automatically calculate the Tukey analysis and graphically display the groups and the statistically significant differences in means. It can then be determined if there are differences in means for groups of programs based on either program office location or

platform type. If it is determined that there is a difference in means for certain groups, then different recommended ranges can be provided to cost estimators based on program type

The final research question is, “Do coefficient of variations for Air Force programs change over time?” A paired t-test is performed to analyze the change in CV over time. The paired t-test is used for comparing the effects in similar samples (Sachs 1982: 308). The samples in this analysis represent the program’s coefficient of variation at the earliest documented point in the program, and the program’s coefficient of variation at the latest documented point in the program. For example, the C-5 RERP contains a CV calculation in FY04 which will represent the earliest calculated CV. The CV calculation for FY10 represents the latest documented calculation. The difference will be the latest calculation less the earliest calculation and represent the column for “Program Office Estimate Coefficient of Variation Change.” The paired t-test will test the significance of the mean being less than zero (Sachs 1982: 308).

$$\begin{aligned} H_o &\geq 0 \\ H_a &< 0 \end{aligned} \tag{3.9}$$

If the mean is less than zero, the test will reject and conclude that the CV does decrease over time for Air Force Programs. It would then be plausible to conclude that different CV ranges be used based on the location of a program in its acquisition lifecycle.

Conclusion

This chapter describes the data used in the analysis and highlights the limitations of the chosen data. There are two forms of data in this study: primary and secondary. The primary data are unique to this study and consist of briefings from program offices. The secondary data are more traditional to other cost growth studies because the data are retrieved from Selective Acquisition Reports. The processes and procedures are described in detail to enable the reader to replicate the analysis. The study employs a Tukey analysis to test for differences in means for particular groupings of programs. If it is determined there is a difference in means of the groups, then appropriate ranges for CV calculations are recommended for each of the groups using Chebshev's rule. Lastly, paired t-tests are used to analyze the differences of means between coefficients of variations to facilitate whether or not CVs decrease over a program's lifecycle.

IV. Results

This chapter presents the results of the three research questions proposed in chapter one: 1) Does analyzing the coefficient of variation ranges from Selective Acquisition Reports and Program Office Estimates match the coefficient of variation ranges provided by AFCAA in the Air Force Cost Risk and Uncertainty Handbook; 2) Should there be different coefficient of variation ranges for dissimilar platform types (aerospace, electronics, and Space and Missiles) for Air Force programs; 3) Do coefficient of variations for Air Force programs change over time? The techniques described in chapter three are utilized to produce the results of the analysis. The relevance of the results is described followed by the accuracy and limitations of the results.

CV Range Benchmarks

The Coefficient of Variation (CV) ranges provided by AFCAA in the AFCRUH are 35-45% for space systems, 25-35% for aerospace systems, and 10-20% for electronic systems (AFCAA, 2007: 8). AFCAA used Selective Acquisition Reports (SARs) to conduct their study. Contrary to the AFCAA study, the results of this analysis are derived through two methodologies to achieve the most accurate results possible. The two methodologies utilize independent data sources: Program Office Estimates (POEs) and SARs. The 30 programs analyzed in this study found contrary results to the AFCAA study regardless of the source of data utilized.

Program Office Estimates

The primary data for this study are POEs. The POEs are produced by sources most familiar with the details of a particular program. The data are separated by milestone location defined by the defense acquisition process. The data used to determine a recommended CV range for programs at Milestone A of the acquisition lifecycle are shown in Table 4.1.

Table 4.1 Program Office CV Data at Milestone A

Program	Program Office	Year	Platform Type	Milestone Location	Lifecycle Stage	Development Office	Program Type	CV @ MS A
CRH (H-47)	ASC	2012	Helicopter	A	SDD	NACA	MDAP	0.16
CRH (H-47)	ASC	2012	Helicopter	A	PD	NACA	MDAP	0.16
CRH (H-71)	ASC	2012	Helicopter	A	SDD	NACA	MDAP	0.27
CRH (H-71)	ASC	2012	Helicopter	A	PD	NACA	MDAP	0.22
B-2 DMS	ASC	2010	Avionics	A	SDD	Program Office	MDAP	0.19
B-2 DMS	ASC	2010	Avionics	A	PD	Program Office	MDAP	0.08
B-2 EHF Inc 2	ASC	2010	Avionics	A	SDD	Program Office	MDAP	0.17
B-2 EHF Inc 2	ASC	2010	Avionics	A	PD	Program Office	MDAP	0.37
B-2 EHF Inc 3	ASC	2009	Avionics	A	SDD	Program Office	MDAP	0.18
AOC Inc 10.2	ESC	2009	Computer Sys	A	SDD	AFCAIG	MAIS	0.35
AOC Inc 10.2	ESC	2009	Computer Sys	A	PD	AFCAIG	MAIS	0.15

There are eleven data points from six different programs analyzed for Milestone A CV calculations. The sample is small which limits the integrity of the conclusions, but the analysis is important to cost estimating. Typical cost growth studies have used SARs as the primary data source. SARs are not required for programs until the program reaches Milestone B. Therefore, previous studies have not been able to provide accurate research on cost growth prior to Milestone B. By utilizing POEs, this research overcomes that limitation.

The empirics of analyzing risk and uncertainty in defense acquisitions leads people to believe programs are very risky early on in the acquisition lifecycle. In fact, a program has not begun engineering, manufacturing, or development if it is still in Milestone A. A program is still conceptual in nature and the required technology to complete the program is still developing. Conventional wisdom is that more risk and uncertainty is added to cost estimates prior to Milestone B since nothing is built yet. This study analyzes the CVs calculated by program offices prior to Milestone B to further the research for this conjecture. The distribution for POE Milestone A coefficient of variation calculations and the results of the Shapiro-Wilk Normality test are shown in Figure 4.1.

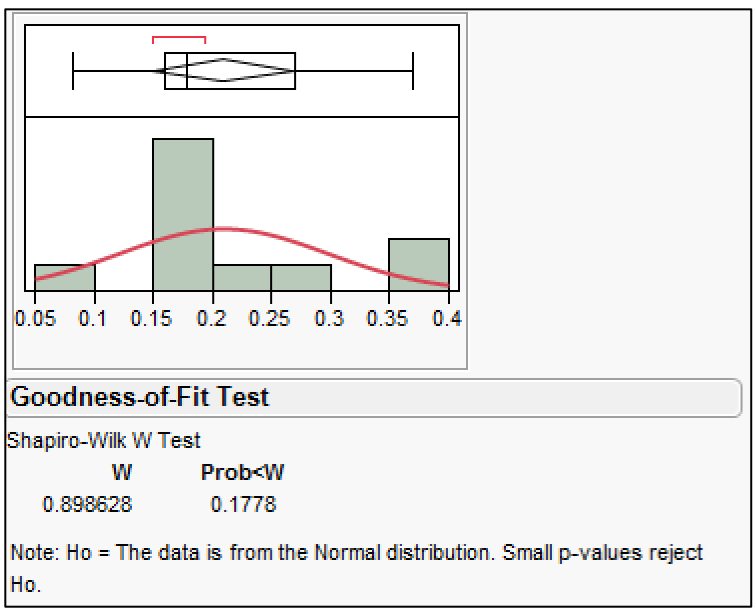


Figure 4.1 CV Distribution for POE at Milestone A

The results show the data are Normally distributed. This is verified by the Shapiro-Wilk test results which has a p-value greater than 0.05, therefore, the null hypothesis fails to reject. Therefore, approximately 99% of the data fall within three standard deviations of the mean. The quantiles of the distribution are shown in Table 4.2.

Table 4.2 CV Quantiles for POE at Milestone A

Quantiles	
100.0%	0.370
99.5%	0.370
97.5%	0.370
90.0%	0.366
75.0%	0.270
50.0%	0.179
25.0%	0.160
10.0%	0.096
2.5%	0.082
0.5%	0.082
0.0%	0.082

The range which encompasses 99% of the data is 0.08 to 0.37. To eliminate outliers and remain consistent with the analysis of all data analyzed in this study, the bottom 25% of the data and top 25% of the data are eliminated to narrow the range to a reasonably accurate recommendation. The middle 50% of the data provide a range of 0.16 to 0.27. This is done because AFCAA currently recommends ranges in 10% intervals and ranges too large provide little insight for decision makers.

The results of the Milestone A analysis show lower than anticipated CV calculations for programs early in the acquisition lifecycle. Regardless of weapon system type, the ranges are below the AFCAA recommendations for program office estimates. Empirically, it is expected, and investigated later in this analysis, that CVs decrease as a

program matures. If that hypothesis holds true and CVs are already below AFCAA recommended ranges, then we expect program office CVs throughout the acquisition lifecycle to be lower than forecasted.

A program enters Milestone B after receiving approval from the Milestone Decision Authority. A program must have mature technology, approved requirements, full funding, approved acquisition strategy, approved acquisition baseline estimate, and an approved contract type (Schwartz, 2013:13). Milestone B is the beginning of developing a physical system. The risk and uncertainty should be closer to the top end of AFCAA recommended ranges because there is little to no actual data to derive a cost estimate. The data used to find a recommended CV range for programs in Milestone B of the acquisition lifecycle are shown in Table 4.3.

Table 4.3 Program Office CV Data at Milestone B

Program	Program Office	Year	Platform Type	Milestone Location	Development Office	Program Type	Lifecycle Stage	CV @ MS B
B-2 EHF Inc 1	ASC	2010	Avionics	B	Program Office	MDAP	SDD	0.17
B-2 EHF Inc 1	ASC	2010	Avionics	B	Program Office	MDAP	PD	0.27
C-5 RERP	ASC	2007	Engine	B	Program Office	MDAP	SDD	0.02
C-5 RERP	ASC	2007	Engine	B	Program Office	MDAP	PD	0.11
CVLSP	ASC	2011	Helicopter	B	Program Office	MDAP	SDD	0.11
CVLSP	ASC	2011	Helicopter	B	Program Office	MDAP	PD	0.08
LAIRCM NexGen MWS	ASC	2012	Electronic	B	Program Office	MDAP	PD	0.03
MQ-1C Gray Eagle	ASC	2009	UAV	B	Program Office	MDAP	SDD	0.15
3 Dim Lng Rng Radar	ESC	2012	Electronic	B	Program Office	MDAP	SDD	0.31
3 Dim Lng Rng Radar	ESC	2012	Electronic	B	Program Office	MDAP	PD	0.27
AFNet Inc 1	ESC	2011	Computer Sys	B	AFCAIG	MAIS	PD	0.02
ITS Inc 2	ESC	2011	Computer Sys	B	AFCAIG	MAIS	PD	0.04
GPS III	SMC	2010	Satellite	B	Program Office	MDAP	SDD&PD	0.18
SBSS Block 10	SMC	2010	Satellite	B	Program Office	MDAP	SDD	0.15

In this analysis ten programs provide fourteen data points. A distribution is fit to the data and analyzed to provide recommendations for CV ranges for programs at Milestone B.

The distribution of the analysis is shown in Figure 4.2.

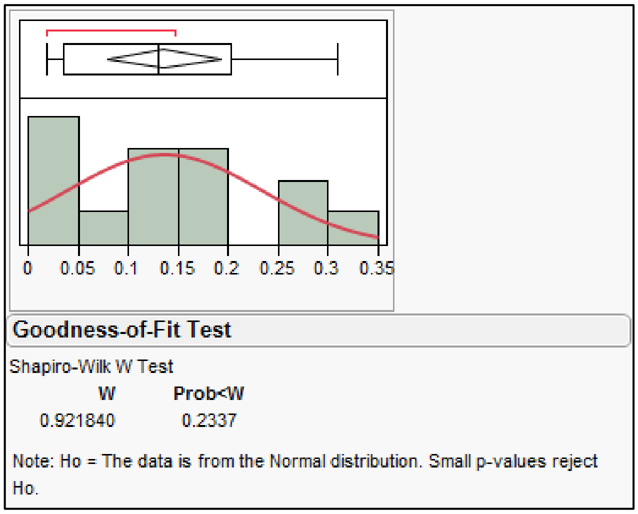


Figure 4.2 CV Ranges Data Analysis for POE at Milestone B

The distribution is Normally distributed with a p-value of 0.2337 which fails to reject the null hypothesis. The quantiles of the distribution are shown in Table 4.4.

Table 4.4 CV Quantiles for POE at Milestone B

Quantiles	
100.0%	0.310
99.5%	0.310
97.5%	0.310
90.0%	0.290
75.0%	0.203
50.0%	0.129
25.0%	0.036
10.0%	0.021
2.5%	0.019
0.5%	0.019
0.0%	0.019

Since the data pass the Shapiro-WIlk test, approximately 99% of the data lie within three standard deviations. The range encompassing 99% of the data is 0.02 to 0.31. However, the range is large and provides little value to decision makers. The middle 50% of the

data show a range of 0.04 – 0.20. This range provides insight and is easier to comprehend for decision makers. The data show that for cost estimates to include a typical amount of risk and uncertainty for programs at Milestone B the CV calculation should be between 0.04 – 0.20. This range is lower than the AFCAA recommendations for CV of 0.10-0.45 which varies depending on weapon system type.

A program must receive permission from the Milestone Decision Authority to enter into Milestone C, Production and Deployment. The programs must have passed developmental testing and operational assessments, demonstrate interoperability, prove affordability, and be fully funded. Entering Milestone C is the beginning of low-rate initial production. If the program passes operational test and evaluation then it can enter into full-rate production (Schwartz, 2013: 10).

Milestone C coefficient of variation calculations are hypothesized to be lower than Milestone A and B because more actual data has been recorded and there are less unknowns and changes in the program. However, changes do occur to programs late in the acquisition lifecycle and the empirics show changes later in the lifecycle cost more than changes earlier in the acquisition lifecycle.

The data used to find a recommended CV range for program in Milestone C of the acquisition lifecycle are shown in Table 4.5.

Table 4.5 Program Office CV Data at Milestone C

Program	Program Office	Year	Platform Type	Milestone Location	Development Office	Program Type	Lifecycle Stage	CV @ MS C
JASSM-ER	AAC	2011	Missile	C	AFCAIG	MDAP	PD	0.20
B-2 EHF Inc 1	ASC	2011	Avionics	C	Program Office	MDAP	SDD	0.08
B-2 EHF Inc 1	ASC	2012	Avionics	C	AFCAIG	MDAP	PD	0.09
C-5 RERP	ASC	2010	Engine	C	AFCAIG	MDAP	SDD	0.11
C-5 RERP	ASC	2010	Engine	C	AFCAIG	MDAP	PD	0.02
C-27J	ASC	2011	Plane	C	Program Office	MDAP	PD	0.14
C-130J	ASC	2011	Plane	C	Program Office	MDAP	PD	0.05
HCMC 130J	ASC	2011	Plane	C	Program Office	MDAP	SDD	0.18
HCMC 130J	ASC	2011	Plane	C	Program Office	MDAP	PD	0.04
MQ-9 Reaper	ASC	2011	UAV	C	AFCAIG	MDAP	SDD	0.14
MQ-9 Reaper	ASC	2012	UAV	C	AFCAIG	MDAP	PD	0.13
MQ-1C Gray Eagle	ASC	2011	UAV	C	Program Office	MDAP	SDD	0.25
MPS Inc III	ESC	2009	Computer Sys	C	AFCAIG	MAIS	PD	0.21
MPS Inc IV	ESC	2010	Computer Sys	C	AFCAIG	MAIS	LCC	0.27

The sample of POEs from Milestone C includes fourteen data points from ten programs. A distribution is fit to the data points and analyzed to provide a range for CV at Milestone C. The result of the fitted distribution for POE calculated CVS at Milestone C is shown in Figure 4.3.

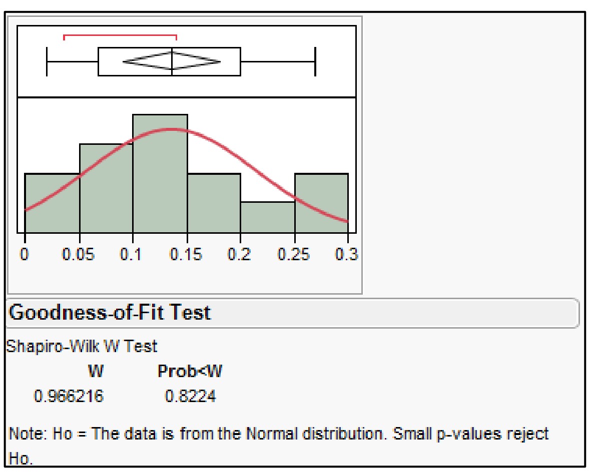


Figure 4.3 CV Ranges Data Analysis for POE at Milestone C

The results of the data analysis pass the Shapiro-Wilk test and represent a population that is Normally distributed. Therefore, 99% of the data lie within three standard deviations of the mean. The 99% range is 0.02 – 0.27. The middle 50% of the data fall between 0.07 – 0.20. The quantiles of the analysis are shown in Table 4.6.

Table 4.6 CV Quantiles for POE at Milestone C

Quantiles	
C.I.	CV
100.0%	0.270
99.5%	0.270
97.5%	0.270
90.0%	0.260
75.0%	0.199
50.0%	0.136
25.0%	0.069
10.0%	0.028
2.5%	0.020
0.5%	0.020
0.0%	0.020

Milestone C marks the beginning of production where actual data is collected which aides cost estimators with forecasting future expenses. Cost estimators are able to collect actual data and have a greater understanding of the program requirements. Therefore, Milestone C coefficient of variations might be lower than Milestone A and B and possibly the recommended ranges by AFCAA.

The analysis of Program Office Estimates CV calculations are summarized in Table 4.7. The current AFCAA recommendations are shown for comparison reasons in Table 4.8.

Table 4.7 AFIT Study POE CV Ranges by Milestone Location

AFIT Study		
A	B	C
16-27%	4-20%	7-20%

Table 4.8 AFCAA Recommend CV Ranges by Weapon System Type

AFCAA		
Electronics	Aerospace	Space
10-20%	25-35%	35-45%

The results aid decision makers with assessing the validity of the cost estimates. If a cost estimate falls outside of the calculated Program Office CV ranges than a decision maker should take a deeper look into the procedures and methods used to derive the cost estimate. If the CV falls outside of the CV ranges calculated from the 99% confidence intervals then a decision maker should seriously question the validity of the POE.

The results depict a more serious concern that POEs do not include enough risk and uncertainty. The CV ranges calculated from the program office data are lower than the ranges recommended by AFCAA in the Air Force Cost Risk and Uncertainty Handbook. With numerous studies reviewed in Chapter 2 where cost growth historically averages between 46-60%, conventional wisdom indicates more risk and uncertainty would be added in cost estimates than the AFCAA recommended values, not less.

Selective Acquisition Reports

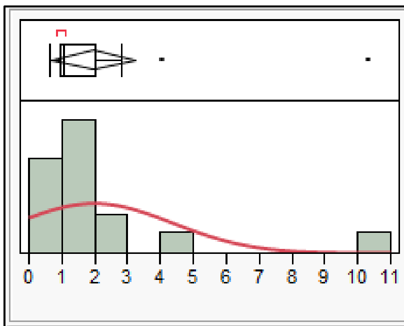
The next step in this analysis is analyzing Selective Acquisition Reports to gain a broader understanding of CV ranges for cost estimates. The SAR estimates for this study are separated into Milestone B and C. There are no Milestone A calculations because SARs are not required for programs prior Milestone B. The SARs are used to calculate the Cost Growth Factor (CGF). A distribution is fit to the CGF data and analyzed. The recommendations for the SAR data utilize Chebyshev's Theorem and the middle fifty-percent of the sample described in Chapter three, because the data does not pass the Shapiro-Wilk test and is therefore not from a Normal distribution. Chebyshev's Theorem allows the analyst to recommend CV ranges regardless of the shape of the CGF distribution. Chebyshev's Theorem states 89% of the data fall within three standard deviations of the mean of any distribution; however, this wide range does not provide much insight for decision makers. Using the middle fifty-percent of the data narrows the range and provides decision makers with a reasonable range to evaluate risk and uncertainty in cost estimates

The data used to find a recommended CV range using SAR data for programs in Milestone B of the acquisition life are shown in Table 4.9.

Table 4.9 SAR CGF Data at Milestone B

Program	Program Office	Year	Platform Type	Milestone Location	Program Type	PHASE	CGF @ MS B
B-2 EHF Inc 1	ASC	2010	Avionics	B	MDAP	SDD	1.03
B-2 EHF Inc 1	ASC	2010	Avionics	B	MDAP	PD	0.85
C-5 RERP	ASC	2007	Engine	B	MDAP	SDD	1.00
C-5 RERP	ASC	2007	Engine	B	MDAP	PD	0.96
C-130 AMP	ASC	2001	Avionics	B	MDAP	PD	1.05
Global Hawk	ASC	2009	UAV	B	MDAP	SDD	2.23
Global Hawk	ASC	2009	UAV	B	MDAP	PD	4.03
MQ-9 Reaper	ASC	2009	UAV	B	MDAP	SDD	1.00
MQ-9 Reaper	ASC	2009	UAV	B	MDAP	PD	1.00
MQ-1C Gray Eagle	ASC	2009	UAV	B	MDAP	SDD	1.30
AFNet Inc 1	ESC	2011	Computer Sys	B	MAIS	PD	0.96
ITS Inc 2	ESC	2011	Computer Sys	B	MAIS	PD	0.66
GPS III	SMC	2010	Satellite	B	MDAP	SDD&PD	1.33
SBIRS SAR	SMC	2011	Satellite	B	MDAP	SDD	2.83
SBIRS SAR	SMC	2011	Satellite	B	MDAP	PD	10.32
SBSS Block 10	SMC	2010	Satellite	B	MDAP	SDD	1.11

A distribution is fit to the data to capture recommended CV ranges representative of the sample. The results of the analysis are shown in Figure 4.4.



Goodness-of-Fit Test

Shapiro-Wilk W Test

W	Prob<W
0.527387	<.0001*

Note: Ho = The data is from the Normal distribution. Small p-values reject Ho.

Figure 4.4 CGF Data from SARs at Milestone B

The initial analysis utilizes sixteen data points from eleven programs. The data are not from the Normal distribution and are skewed right. The sample is heavily influenced by extreme cost growth outliers. The quantity and requirements of the SBIRs program has led to extreme cost growth of the program. Also, the Global Hawk UAV has proven extremely useful in wartime theatre. The users have demanded more Global Hawks with better technology. The result has been cost growth greater than 300% of the initial baseline.

The results show that 89% of the data lie between 0.23 –3.62. The middle fifty percent fall between 1.20 – 2.48. Although some cost growth is expected because of quantity and requirements changes, as discussed in chapter 2, the production cost growth associated with SBIRs and Global Hawk are extreme and atypical. The quantiles of the analysis are shown in Table 4.10.

Table 4.10 CV Quantiles for SARs at Milestone B

Quantiles		
C.I.	CGF	CV
100.0%	10.323	3.623
99.5%	10.323	3.623
97.5%	10.323	3.623
90.0%	5.916	3.021
75.0%	2.003	2.475
50.0%	1.039	2.305
25.0%	0.967	1.195
10.0%	0.792	0.405
2.5%	0.661	0.232
0.5%	0.661	0.232
0.0%	0.661	0.232

The outliers' effects on the sample lead to an analysis which removes the data points that can be explained by extreme quantity and requirements increases. The result

of the distribution analysis after the production CGFs for SBIRs and Global Hawk are removed is shown in Figure 4.5.

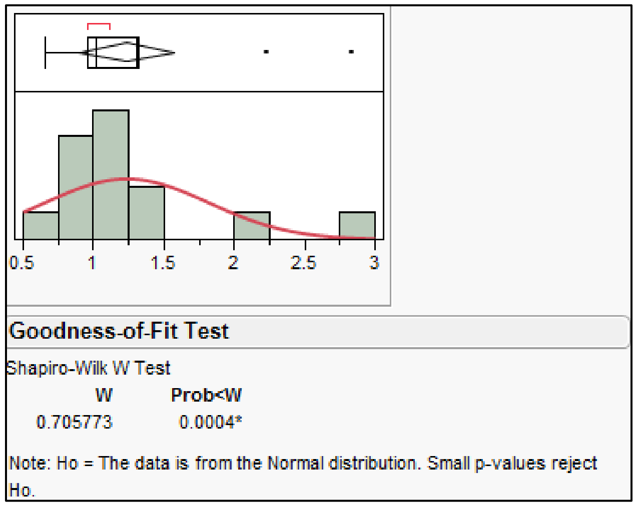


Figure 4.5 CGF Data from SARs at Milestone B Outliers Removed

The analysis shows the sample is not Normally distributed because the p-value for the Shapiro-Wilk test is 0.0004. The CV recommendations for Milestone B after the outliers are removed shows 89% of the data fall between 0.20 - 0.88. Since the 89% range is extremely large, the recommendation is narrowed to the middle 50% of the data. The results are CVs between 0.45 - 0.61.

The quantiles of the distribution are shown in Table 4.11.

Table 4.11 CV Quantiles for SARs at Milestone B Outliers Removed

Quantiles		
C.I.	CGF	CV
100.0%	2.825	0.882
99.5%	2.825	0.882
97.5%	2.825	0.882
90.0%	2.527	0.773
75.0%	1.309	0.610
50.0%	1.017	0.574
25.0%	0.957	0.445
10.0%	0.755	0.231
2.5%	0.661	0.206
0.5%	0.661	0.206
0.0%	0.661	0.206

The CV recommendation aides decision makers with determining if enough risk and uncertainty is built into a cost estimate. The Milestone B CGF analysis shows decision makers an estimate with a CV between 45-61% is consistent with programs in Milestone B.

The Milestone C CGF analysis is conducted using the same methods to determine the Milestone C ranges. The data used to find a recommended CV range using SAR data for programs in Milestone C of the acquisition life are shown in Table 4.12.

Table 4.12 SAR CGF Data at Milestone C

Program	Program Office	Year	Platform Type	Milestone Location	Program Type	PHASE	CGF @ MS C
JASSM-ER	AAC	2011	Missile	C	MDAP	PD	1.43
B-2 EHF Inc 1	ASC	2011	Avionics	C	MDAP	PD	1.10
B-2 EHF Inc 1	ASC	2012	Avionics	C	MDAP	SDD	0.77
C-5 RERP	ASC	2010	Engine	C	MDAP	SDD	0.98
C-5 RERP	ASC	2010	Engine	C	MDAP	PD	1.00
C-27J	ASC	2011	Plane	C	MDAP	PD	0.56
C-130 AMP	ASC	2011	Avionics	C	MDAP	PD	0.11
C-130J	ASC	2011	Plane	C	MDAP	SDD	36.09
C-130J	ASC	2011	Plane	C	MDAP	PD	16.38
HCMC 130J	ASC	2011	Plane	C	MDAP	SDD	1.59
HCMC 130J	ASC	2011	Plane	C	MDAP	PD	1.04
MQ-9 Reaper	ASC	2011	UAV	C	MDAP	SDD	1.29
MQ-9 Reaper	ASC	2011	UAV	C	MDAP	PD	1.06
MQ-1C Gray Eagle	ASC	2011	UAV	C	MDAP	SDD	1.04
MPS Inc III	ESC	2009	Computer Sys	C	MAIS	PD	1.07
MPS Inc IV	ESC	2010	Computer Sys	C	MAIS	PD	0.74

Again, a distribution is fit to the data to analyze appropriate CV ranges for MDAPs at Milestone C. The result of the Milestone C CGF analysis is shown in Figure 4.6.

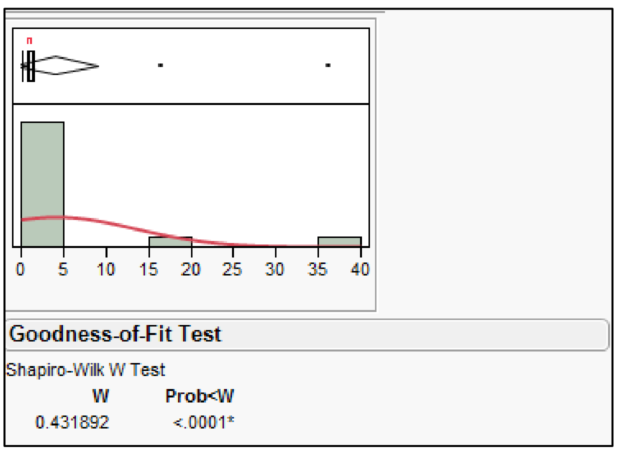


Figure 4.6 CGF Data from SARs at Milestone C

The CV recommendations for Milestone C shows 89% of the data fall between 0.26 - 86.73. The recommendation is narrowed to the middle 50% of the data which yields CV results between 6.71 – 11.34. The recommend range is extremely large because outliers from the C-130J and C-130 AMP programs. The C-130J has increased the quantity to be purchased by 1500% from the original contract. The C-130 AMP was cancelled and has a CGF that is atypical for that reason. The quantiles for the anlaysis are shown in Table 4.13.

Table 4.13 CV Quantiles for SARs at Milestone C

Quantiles		
C.I.	CGF	CV
100.0%	36.090	86.727
99.5%	36.090	86.727
97.5%	36.090	86.727
90.0%	22.292	22.102
75.0%	1.393	11.338
50.0%	1.048	8.920
25.0%	0.825	6.714
10.0%	0.423	0.420
2.5%	0.108	0.259
0.5%	0.108	0.259
0.0%	0.108	0.259

To account for the outliers, the analysis is performed again with the data points removed from the sample.

A distribution is fit to the data and the results are analyzed. Figure 4.7 shows the results after the outliers for the C-130J and C-130 AMP programs are removed.

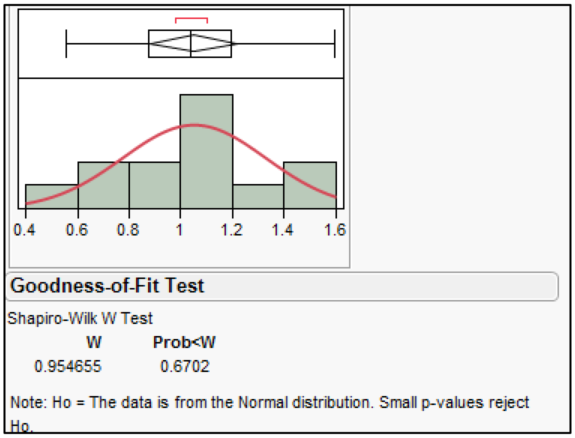


Figure 4.7 CGF Data from SARs at Milestone C with Outliers Removed

The Milestone C CGF analysis with outliers removed represents a Normal distribution and shows 99% of the data are between 0.17-0.50. The middle 50% of the data narrow the range to 0.23 – 0.32.

Table 4.14 CV Quantiles for SARs at Milestone C Outliers Removed

Quantiles		
C.I.	CGF	CV
100.0%	1.593	0.496
99.5%	1.593	0.496
97.5%	1.593	0.496
90.0%	1.527	0.438
75.0%	1.195	0.316
50.0%	1.038	0.267
25.0%	0.876	0.232
10.0%	0.632	0.181
2.5%	0.558	0.174
0.5%	0.558	0.174
0.0%	0.558	0.174

The results of the CGF coefficient of variation analysis at Milestone B and C calculated using Selective Acquisition Reports are shown in Table 4.15. For comparative reasons, the results of the NCCA are shown in Table 4.16 (Flynn, 2011:30).

Table 4.15 Air Force SAR Calculated CV Ranges by Milestone Location

Air Force SAR Data		
Milestone A	Milestone B	Milestone C
N/A	45-61%	23-32%

Table 4.16 NCCA Calculated CV Ranges by Milestone Location

NCCA Study		
Milestone A	Milestone B	Milestone C
41-74%	31-54%	21-34%

These results aid decision makers in gauging whether a cost estimate is built with enough risk and uncertainty. The SAR data shows that a cost estimate with a CV within one of the ranges shown in Table 4.14 is built with a typical amount of risk and uncertainty based on historical programs.

The combination of the Program Office calculated CVs and the SAR calculated CVs provides further insight into appropriate CV benchmarks. The results of POE and SAR calculated CVs are shown in Table 4.17.

Table 4.17 POE and SAR Calculated CV Ranges by Milestone Location

SAR All Data			
	Milestone A	Milestone B	Milestone C
POE	16-27%	4-20%	7-20%
SAR	N/A	45-61%	23-32%

The ranges should be used to determine if there is enough risk and uncertainty in cost estimates. The results of the study are intriguing. Ideally, the recommend CV ranges should be the same. The facts are that the CV ranges from SAR and POE calculations are drastically different. Several arguments could be made for either approach. The POE CVs could be based on risk adjusted estimates that are much higher than the programmed amount shown in SARs. Less risk and uncertainty would be built into estimates that have higher values; further research is needed to investigate this possibility. The POE CVs could also be lower because of pressures to secure funding or over overoptimistic assumptions that do not formulate. To further complicate matters, the AFCAA ranges fall somewhere between the POE and SAR calculated CVs. The objective of this study is not to state which range is superior compared to the other. The goal is to further the research of the usefulness of the coefficient of variation in the cost estimating career field and provide rigor to recommended ranges.

CV by Platform Type

The Air Force Cost Analysis Agency recommends ranges based on the type of program. The research by Brain Flynn and Paul Garvey for the Naval Center for Cost Analysis shows for Navy programs that there is no statistical difference that suggests there should be a different CV due to the type of weapon system developed. In order to further the research of this topic, this study analyzes whether or not there is a statistical difference between the type of weapon system and the amount of risk and uncertainty in Air Force programs. This study analyzes the POE calculated and SAR calculated CVs

using the Tukey Test outlined in chapter 3 to investigate whether or not CV ranges should be different because of the type of weapon system.

Relating back to the original research questions, the objective of the study is to investigate if the CV should be different based on the type of weapon system. The research aims to collect enough data to compare the CV by the type of weapon system: helicopter, plane, computer system, satellite, UAV, electronics, or missile. However, due to the limited amount of data available for this research it is not possible to get a statistically significant result based on the stated criteria. Nevertheless, there is enough data to analyze the platform type by separating the programs by the program office location: Aerospace Systems Center (ASC), Electronics System Center (ESC), and Space and Missile Center (SMC). The product centers represent the different weapon system platform types: ASC represents aircraft, ESC represents electronics, SMC represents space systems.

Platform Type: POE

The program offices' calculated CVs are compared to distinguish differences in CV ranges based on weapon type which is determined by the product center location. The mean of all the CVs for a product center are compared against the other product centers in the sample using the Tukey test which is a means comparison test. The null hypothesis of the Tukey test is that all the means are equal.

$$\begin{aligned} H_0 &= \mu_i = \mu_j \\ H_1 &= \mu_i \neq \mu_j \end{aligned} \tag{4.1}$$

The results of the POE calculated CVs for Milestone A are shown in Table 4.18.

Table 4.18 Tukey Analysis Results POE CVs at Milestone A

POE MS A Means Comparison(P-Values)		
Platform Type	Electronics	Aircraft
Electronics		0.7069
Aircraft		

With the p-value of 0.7069, the null hypothesis fails to reject and the conclusion is that all the means are equal. The results of the Tukey test show that at the alpha level of 0.05 there are no statistically significant differences between CVs of programs developed at Electronics Systems Center and Aerospace Systems Center. Eleven programs data points representing six programs are analyzed. There is limited data for programs in Milestone A, but the finding is the beginning of a trend when evaluating CV differences among varying program types.

The study also looked at differences among program types at Milestone B. The results of the analysis are shown in Table 4.19.

Table 4.19 Tukey Analysis Results POE CVs at Milestone B

POE MS B Means Comparison(P-Values)			
Platform Type	Electronics	Aircraft	Space
Electronics		0.7696	0.9967
Aircraft			0.8092
Space			

The results show that at an alpha level of 0.05 there are no statistically significant similarities between CVs based on program type. The analysis looks at 14 data points among ten programs developed at the Aerospace Systems Center, Electronics System

Center, and Space and Missile Systems Center. The result is consistent with the results found at Milestone A.

The data are then analyzed at Milestone C. The results of the Milestone C analysis are shown in Table 4.20.

Table 4.20 Tukey Analysis Results POE CVs at Milestone C

POE MS C Means Comparison (P-Values)		
Platform Type	Electronics	Aircraft
Electronics		0.0669
Aircraft		

The trend remains constant as there is no statistically significant difference at an alpha level of 0.05 between CV ranges based on program type. The Milestone C analysis includes thirteen data points from nine different programs from the Electronics Systems Center and Aerospace Systems Center.

The significant finding from this research is that there should not be different CV ranges based on product center or program type. This finding is contrary to the AFCAA recommendation in the Air Force Cost Risk and Uncertainty Handbook. The popular belief among cost estimators is that different types of weapon systems have different levels of risk and uncertainty. Brian Flynn and Paul Garvey spearheaded the analysis for naval programs and found there is no significant difference among platform types for Navy programs. The research conducted for this study supports Flynn and Garvey's findings, but for Air Force programs, when using the POE calculated CVs.

Platform Type: SARs

The SARs Cost Growth Factors are analyzed to provide more integrity to coefficient of variation studies for DoD programs. The study compares CGFs based on product center location to determine if there are significant differences among CGFs. If there are differences in CGFs, then it is plausible that there should be different CV ranges based on weapon system platform type. The Cost Growth Factors for each program are analyzed using the Tukey test to compare means of the varying platform types. The results of the CGF analysis for weapon systems at Milestone B are shown in Table 4.22.

Table 4.21 Tukey Analysis Results SARs at Milestone B

SAR MS B Means Comparison(P-Values)			
Platform Type	Electronics	Aircraft	Space
Electronics		0.9295	0.2857
Aircraft			0.1941
Space			

The results show there is no statistically significant evidence to state the CVs should be different based on program type at the alpha level of 0.05 using CGFs from SARs. The data include sixteen points from eleven programs. The SBIRS and Global Hawk outliers discussed earlier in the chapter are included in the analysis. The results of the analysis with the SBIRS and Global Hawk removed are shown in Table 4.22.

Table 4.22 Tukey Analysis Results SARs at Milestone B Outliers Removed

SAR MS B Means Comparison(P-Values)			
Platform Type	Electronics	Aircraft	Space
Electronics		0.6940	0.1776
Aircraft			0.2611
Space			

The SARs are also analyzed at Milestone C to determine if there is a difference between program types later in the acquisition lifecycle. The results are summarized in Table 4.23.

Table 4.23 Tukey Analysis Results SARs at Milestone C

SAR MS C Means Comparison(P-Values)		
Platform Type	Electronics	Aircraft
Electronics		0.1941
Aircraft		

The results of the analysis show that there are no differences between platform type based on the SAR CGF calculations. The analysis shown in Table 4.22 includes the C-130J and C-130 AMP outlier programs that were removed earlier when answering the first research questions of this analysis. As a reminder, the outliers are removed in the next step of the analysis because of drastic quantity increases for the C-130J and program cancellation for the C-130 AMP.

In order to remain consistent throughout the analysis, the SAR CGF comparisons are analyzed with the outliers removed. The results are shown in Table 4.24.

Table 4.24 Tukey Analysis Results SARs at Milestone C Outliers Removed

SAR MS C Means Comparison(P-Values)		
Platform Type	Electronics	Aircraft
Electronics		0.5496
Aircraft		

The results again yield no statistical difference between program types based on SAR CGF calculations.

The results of this analysis are consistent with the POE calculated CVs and the Naval Center for Cost Analysis (NCCA) calculated CVs by Flynn and Garvey. However, the results differ from the recommended ranges from AFCAA which specify ranges based on platform differentiated by product center. The lack of insight to the AFCAA study and the repeated finding that there are no statistically significant relationships between program type and CV range leads to the conclusion that there should not be different CV ranges based on program type.

The sample used for this analysis is not large enough to compare every program office at every milestone point. The results are only useful for the comparisons between the program offices within the sample. However, the results remain the same regardless of which CV calculation is used and regardless of milestone which provides strong evidence to support the claim that there is no value in recommending CV ranges based on platform type. Further research with a larger data set would provide integrity to the results of the analysis. In addition, the study attempted to compare CV ranges based on weapon system type: helicopters, airplanes, missiles, UAVs, electronics, and avionics, but the sample is not large enough to draw statistical inferences by weapon system categories.

CV Changes Over Time

The final research question in this analysis is determining whether CV ranges should change as a weapon system progresses through the acquisition lifecycle. The Air

Force Cost Risk and Uncertainty Handbook recommends one CV range throughout the acquisition lifecycle. Contrarily, the NCCA study found that CVs decrease overtime. The objective of this study is to further the research of appropriate CV ranges for DoD programs by analyzing Air Force programs through Program Office Estimates and Selective Acquisition Reports.

Changes Over Time: POEs

The study uses a paired t-test to determine if CVs decrease over the acquisition lifecycle. The first calculated CV from programs are subtracted from the last calculated CV for the program. The selection criteria for the data are that there must be two CV calculations for the same program within System Development and Demonstration or Production and Deployment lifecycle stages. The paired t-test requires the differences to be Normally distributed for the test to be valid. The data are shown in Table 4.25.

Table 4.25 POE Data for Changes on CV in Time

Program	Program Office	Year	Platform Type	Development Office	Program Type	Program Phase	Last CV - First CV
JASSM-ER	AAC	2011	Missile	AFCAIG	MDAP	PD	-0.05
B-2 EHF Inc 1	ASC	2011	Avionics	Program Office	MDAP	SDD	-0.20
B-2 EHF Inc 1	ASC	2012	Avionics	AFCAIG	MDAP	PD	-0.08
C-5 RERP	ASC	2010	Engine	AFCAIG	MDAP	SDD	0.09
C-5 RERP	ASC	2010	Engine	AFCAIG	MDAP	PD	-0.09
C-27J	ASC	2011	Plane	Program Office	MDAP	PD	0.13
HCMC 130J	ASC	2011	Plane	Program Office	MDAP	SDD	-0.17
HCMC 130J	ASC	2011	Plane	Program Office	MDAP	PD	-0.01
MQ-9 Reaper	ASC	2012	UAV	AFCAIG	MDAP	SDD	0.02
B-2 EHF Inc 2	ASC	2010	Avionics	Program Office	MDAP	SDD	-0.02
MQ-1C Gray Eagle	ASC	2011	UAV	Program Office	MDAP	SDD	0.10
3 Dim Lng Rng Radar	ESC	2012	Electronic	Program Office	MDAP	SDD	-0.09
3 Dim Lng Rng Radar	ESC	2012	Electronic	Program Office	MDAP	PD	-0.03
MPS Inc IV	ESC	2010	Computer Sys	AFCAIG	MAIS	LCC	0.03
GPS III	SMC	2010	Satellite	Program Office	MDAP	LCC	-0.07
SBIRS GEO 1-2	SMC	2011	Satellite	Program Office	MDAP	SDD	0.00
SBIRS SFP GEO 3	SMC	2010	Satellite	Program Office	MDAP	SDD	-0.01
SBIRS SFP GEO 4	SMC	2010	Satellite	Program Office	MDAP	SDD	0.00
SBSS Block 10	SMC	2010	Satellite	Program Office	MDAP	SDD	-0.11

The analysis includes nineteen data points from fifteen programs. The hypothesis of the analysis is that CVs will decrease over time because as a program matures there are more actual data which aid cost estimators with assessing risk and uncertainty. The results of the analysis are depicted in Figure 4.8 and Figure 4.9.

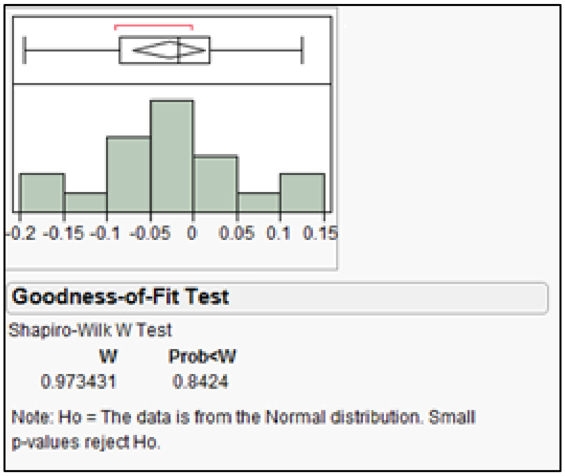


Figure 4.8 Paired T-Test Distribution of POE Calculated CVs

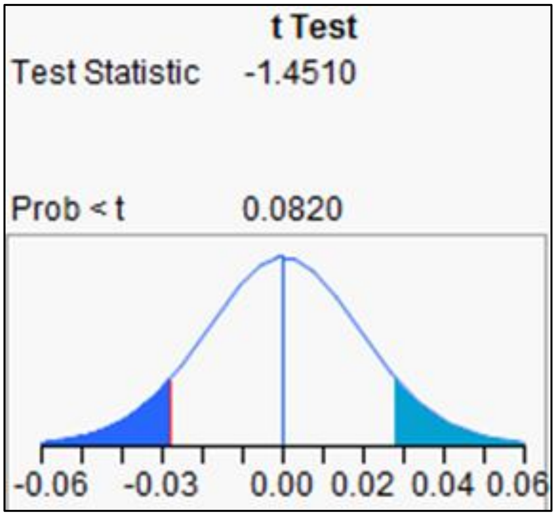


Figure 4.9 Paired T-Test P-Value

The null and alternative hypotheses of this t-test are shown in equation 4.2.

$$\begin{aligned} H_o &\geq 0 \\ H_a &< 0 \end{aligned} \tag{4.2}$$

The results show the paired t-test is not statistically significant at an alpha level of 0.05. However, the results are significant at an alpha level of 0.10. The data pass the Shapiro-Wilk test validating the assumption of normality. The results do not overwhelmingly state that CVs decrease over time based on this analysis, but it does not drastically state that CVs do not decrease overtime. In statistical terms, the conclusions drawn from the test could violate a Type I or Type II error because the results are close to the critical value. A Type I error is an incorrect rejection of the null hypothesis. For this analysis, Type I error represents stating the CVs decrease over time when they in fact do not decrease as a program matures. A Type II error is failing to reject a false null hypothesis. By stating that the CVs do not decrease as a program matures after analyzing this data, the drawn conclusions are in jeopardy of violating a Type II error because the null hypothesis fails to reject; however, if the null is in fact true that the CVs do decrease over time then the analysis has committed a Type II error.

Changes Over Time: SARs

To further the research on whether or not CVs decrease over time, the CGFs are compared between Milestone B and C calculated earlier. A paired t-test is not appropriate for this analysis because there is only one CGF calculation per program. The POE analysis utilizes two CV calculations per program.

The results of the CGF range recommendations are again depicted in Table 4.26 below.

Air Force SAR Data		
Milestone A	Milestone B	Milestone C
N/A	45-61%	23-32%

Table 4.26 SAR Calculated CV Ranges by Milestone Location

A comparative analysis shows CVs do decrease over time based on the CVs calculated from the Cost Growth Factors.

To summarize, the conclusion as to whether or not CVs decrease over time is that there is statistical evidence which supports a decreasing CV range as a program matures through its lifecycle. The conclusion is based on choosing a possible Type I error of stating the CVs decrease as a program matures when, in fact, they do not decrease. The decision to draw the stated conclusion is based on the supporting Cost Growth Factor analysis which clearly shows a decrease in CV over time. The evidence is not as powerful as hypothesized, but the POE and SAR analysis of this study combined with the NCCA study supports the relationship of CVs decreasing over time.

Limitations

The results of the study have limitations based on the quantity and quality of data. This is the first study analyzing coefficient of variation using source data from program office cost estimating briefings. The availability of data is constrained to the number of records archived by program offices. There is not an Air Force Instruction requiring program offices to maintain risk analysis data. This limits the number of programs that are analyzed. It would be beneficial for future research to require program offices or the

SARs to maintain the CV of the estimates. The smaller quantity of programs analyzed decreases the power of the statistical tests and decreases the certainty in the results. Ideally, the analysis would use annual AFCAIG briefs for every year of the program since the Analysis of Alternatives segment of the acquisition lifecycle.

This study employs thirty Air Force ACAT I programs. The programs are not evenly distributed amongst the four different product centers. The programs are also not evenly distributed among the different acquisition lifecycle milestones. Ideally, it is beneficial to have a similar number of programs from each product center representative of the population of ACAT I programs. However, this research is sponsored by the former Aerospace Systems Center (ASC), today known as Life Cycle Management Center. The data is provided primarily by ASC which skews the data towards programs related to aircraft.

The cost growth studies reviewed in the literature review vary on the methodology for determining the Cost Growth Factor. There are studies that normalize the CGF for quantity increases and decreases by dividing the estimate by the current quantity stated in the SAR. The studies normalize for the quantity in order to analyze the data in terms of cost growth per unit. A limitation of this analysis is that it does not analyze cost growth per unit. This study analyzes cost growth on the program holistically. The reason this study does not analyze cost growth per unit is because decision makers decide to enter Milestone B of the acquisition phase because they are under the assumption that they can procure a specific quantity at a specific cost. It is the cost estimator’s responsibility to forecast, as accurately as possible, a realistic procurement quantity for the resources invested.

Summary

The objective of this chapter is to explain the results of the study using the methodology defined in chapter three to answer the research questions proposed in chapter one. The goal of the first research question is to validate the CV ranges from the Air Force Cost Risk and Uncertainty Handbook through the analysis of POE and SAR calculated CVs. The research found different CV ranges than those recommended in the Air Force Cost Risk and Uncertainty Handbook. The research also found different CV ranges based on the type of data analyzed. The conclusion furthers the research into appropriate CV ranges for Air Force Major Defense Acquisition Programs.

The objective of the second research questions is to analyze whether CV ranges should be different based on the type of weapon system analyzed. The Air Force Cost Risk and Uncertainty Handbook recommends different ranges based on platform type. This research analyzed the platform type using POE and SAR calculated CVs. The results show there are no statistically significant differences between Air Force platform types and coefficient of variation. The results coincide with research performed by NCCA on Navy programs in 2012 which also found no difference in CV based on various weapon system platforms.

The last research question explores the notion that CVs should decrease as a program matures through the acquisition lifecycle. The Air Force Cost Risk and Uncertainty Handbook does not provide different CV ranges based on the maturity of the program. The NCCA study found CVs decrease over time. This research used POE and SAR calculated CVs to further the research. The results showed there is statistical evidence which supports CVs decreasing over time, but not to the tested significance

level. However, the SAR calculated CVs clearly depict CVs decrease as a program matures.

The next chapter, the conclusions, summarizes the results of the three research questions proposed in chapter one. It then discusses the implications of the findings for decision makers. Lastly, it highlights topics for follow-on research in risk and uncertainty benchmarks for major defense acquisition programs.

V. Conclusions

The goal of this analysis is to answer the research questions developed in Chapter 1. Simplified, the intentions are to recommend coefficient of variation (CV) ranges for Air Force Acquisition programs, determine if different CV ranges should be used base on platform type, and determine if CV decreases over the course of the program's acquisition lifecycle. This chapter will briefly recap the results of each research questions and make a recommendation for each research question. The implications of the study and the impacts of the recommendation will be discussed. Lastly, potential follow-on research topics will be addressed to conclude this study.

Recommended CV Ranges

The intent of the first research question is to provide Air Force cost estimators with coefficient of variation benchmarks for Air Force weapon systems. The study uses data from Program Office Estimates (POEs) and Selective Acquisition Reports (SARs). This is the first study to analyze cost growth and CV benchmarks utilizing source data from program offices. The results of this study are compared with previous research in the same arena. The Air Force Cost Analysis Agency (AFCAA) and the Naval Center for Cost Analysis (NCCA) performed studies which recommend CV benchmarks for their respective services. The results of both methodologies (POE and SAR) employed in this study are compared with the AFCAA and NCCA studies before making a recommendation.

The results of the studies are shown in Tables 5.1 and 5.2.

Table 5.1 CVs from AFIT and NCCA Studies

AFIT Summary			
	Milestone A	Milestone B	Milestone C
AFIT- POE	16-27%	4-20%	7-20%
AFIT-SAR	N/A	45-61%	23-32%
NCCA	41-74%	31-54%	21-34%

Table 5.2 CVs from AFCAA Study

AFCAA Results		
Electronics	Aerospace	Space
10-20%	25-35%	35-45%

After comparing the data of this study with the results of the AFCAA and NCCA studies, this study recommends Air Force programs use coefficient of variation benchmarks of: 41-74% during Milestone A, 45-61% during Milestone B, and 23-32% during Milestone C. These recommendations are shown in Table 5.3.

Table 5.3 AFIT Study CV Recommendations

AFIT Study CV Benchmarks		
Milestone A	Milestone B	Milestone C
41-74%	45-61%	23-32%

The recommendation for those benchmarks is made because three of the conclusions from the studies support each other. The results of the AFCAA, NCCA, and SAR analysis of this study are fairly similar. It is not coincidence that three studies make fairly analogous conclusions. Also, sound reasoning derived from the results of the POE

analysis at Milestone B suggest that a program estimate with a low range of 4% variability at a stage in the acquisition lifecycle that has yet to manufacture the weapon system is irrational. The extrapolation of learning curve probably has more than 4% variability, and the actual data to develop the learning curve has not been collected when a program is in Milestone B.

The recommendation of a 41-74% CV benchmark at Milestone A is derived from the NCCA study. This recommendation was not substantiated by the results of the analysis in this study; however, the Milestone B and Milestone C recommendations are fairly similar. It is more beneficial to have a recommended range at Milestone A that is based on Navy programs than to not have a recommendation at all. It is assumed that if the recommendations at Milestone B and Milestone C are fairly analogous that Milestone A benchmarks will also be comparable; however, more research is needed to provide integrity to that part of the analysis.

CVs by Platform Type

The intent of the second research question is to answer whether or not different CV ranges should be employed based on the type of weapon system developed. The question is hypothesized because AFCAA recommends different CVs based on weapon system type; however, the NCCA study found CVs should be the same regardless of weapon system type.

This study analyzed POE and SAR CVs and compared the results with the AFCAA and the NCCA results. The results of both the POE and SAR analysis concluded that CVs should be the same regardless of platform type. Since both analyses

validated the results of the NCCA study, the recommendation is that the same CV ranges should be applied to all weapon systems.

Lifecycle CV Changes

The intent of the third research question is to determine whether or not CV ranges should be different based on the location of the program in the defense acquisition lifecycle. The question is hypothesized because AFCAA recommends one CV range regardless of what stage the program is in during the acquisition lifecycle; however, the NCCA study found that CVs decrease over time because more information is learned as a program progresses through its lifecycle which reduces the risk and uncertainty in the cost estimate.

The results of this study found that at the tested alpha level of 0.05 the POE CVs do not decrease over time; however, at an alpha level of 0.10 the POE CVs do decrease over time. The results were not conclusive enough to make a determination based on the results; however, when the SAR results are used in conjunction with the POE the results trend towards the claim that CVs decrease over time. The combination of the POE, SAR, and NCCA results lead to the conclusion that CVs should decrease as a program matures through the acquisition lifecycle. This conclusion is based on the acceptance of a Type I error based on the corroborating evidence. This claim validates the recommended ranges which change based on the phase in the lifecycle of a program.

The last recommendation of this research is that there is extreme value in analyzing the results of the source data. Decision makers should make it mandatory for program offices and independent agencies to maintain and track changes in CV, point

estimates, and risk-adjusted estimates. This can be done by putting a coefficient of variation disclosure requirement on SARs or by making program offices and independent agencies responsible, and accountable through inspections, for archiving annual peer-review and AFCAIG briefings. During the data collection phase of this study, it was apparent there are no guidelines requiring program offices or independent agencies from archiving old source data. The source data is extremely valuable for an in-depth analysis of the requirements, schedule, and cost changes throughout a program's lifecycle.

Implication of Findings

The implication of the findings is important because the results suggest that cost estimators should add more risk and uncertainty into cost estimates to increase the accuracy. The result of added risk and uncertainty is higher risk-adjusted cost estimates. Higher risk-adjusted cost estimates could lead to the funding of less Air Force programs at a time when the nation is facing budget cuts while our nation is fighting a war in the Middle-East. However, if decision makers are serious about reducing cost growth in the DoD acquisition system then they should enhance the review process to ensure appropriate amounts of risk and uncertainty are added to cost estimates.

The implementation of the recommended CV benchmarks will increase the accuracy of cost estimates. The increased accuracy of estimates will increase the confidence of decision makers in the cost estimating community. The CV recommendations will improve Air Force portfolio analysis because decision makers will have more accurate information regarding the resources needed to fund Air Force requirements.

Follow-On Research

This study employs source data for the first time when evaluating CV ranges for DoD systems. The results of the source data compared with the SAR data are drastic. The findings leave the cost estimating community wondering why the results are so different. Future research could help solve this problem. This study does not analyze whether or not higher program office estimates are a reason for the lower CVs. The estimate in the SARs is the number programmed for in the President's Budget (PB); however, it is hypothesized that the number in the PB is not always the same as the number estimated by the cost community. Comparing the source data estimates with the current estimates in the SAR could help answer this question.

The difference in POE and SAR data could also be attributed to optimistic assumptions or pressures to secure funding which lead to cost growth. Could it be possible that there is a correlation between high POE CVs and less cost growth? Future research could use the source data to determine if programs that implemented higher CVs in the POEs had less cost growth in the SARs. Or if the risk-adjusted estimate in the POE is much higher than the SARs, are the lower CV ranges found in the POEs appropriate?

Lastly, the data employed in this analysis is provided primarily from the sponsor of the research, the Aerospace Systems Center now known as the Lifecycle Management Center. The data is heavily influenced by ASC data. Future research could continue the collection and use of POE data which would increase the power of the results and provide further integrity.

Appendix A: Powerpoint® Slide Examples

U.S. AIR FORCE

Top Level Results (Mean)

	BY11$M	TY$M
EMD	$32.6	$33.5
Production	$3,528.0	$4,268.7
MILCON	$301.3	$354.7
O&S	$22,462.9	$35,322.1

Integrity - Service - Excellence

U.S. AIR FORCE

Production Risk Assessment Summary

Monte Carlo Output (BY11$M):

Point Estimate
3,447.6 BY11$M
4,167.6 TY$M
40% CL

Percentiles:	Forecast values
0%	$ 2,088.4
10%	$ 2,777.9
20%	$ 2,877.8
30%	$ 2,946.6
40%	$ 3,009.7
50%	$ 3,068.4
60%	$ 3,128.2
70%	$ 3,193.7
80%	$ 3,267.7
90%	$ 3,376.4

Mean Confidence
3,528.0 BY11$M
4,268.7 TY$M
Mean (51% CL)
8% CV

$ 4,167.6	Point Estimate (TY$M)
$101.2	Risk Estimate (TY$M)
$4,268.7	Mean Confidence Estimate (TY$M)

Integrity - Service - Excellence

Appendix B: SAR Cost and Funding Example

Cost and Funding

Cost Summary

Total Acquisition Cost and Quantity

	BY2000 $M			BY2000 $M	TY $M		
Appropriation	SAR Baseline Dev Est	Current APB Development Objective	Threshold	Current Estimate	SAR Baseline Dev Est	Current APB Development Objective	Current Estimate
RDT&E	1413.9	1280.4	1408.4	1416.4 [1]	1538.5	1369.9	1577.6
Procurement	7381.0	7574.7	8332.2	7061.6	9551.8	9630.7	9553.3
Flyaway	6660.2	--	--	5964.5	8623.4	--	8072.2
Recurring	6626.2	--	--	5964.5	8583.8	--	8072.2
Non Recurring	34.0	--	--	0.0	39.6	--	0.0
Support	720.8	--	--	1097.1	928.4	--	1481.1
Other Support	517.3	--	--	730.9	664.1	--	990.6
Initial Spares	203.5	--	--	366.2	264.3	--	490.5
MILCON	3.1	3.1	3.4	0.0	3.6	3.6	0.0
Acq O&M	0.0	0.0	--	0.0	0.0	0.0	0.0
Total	8798.0	8858.2	N/A	8478.0	11093.9	11004.2	11130.9

[1] APB Breach

This SAR is submitted with cost and funding data based on the Fiscal Year (FY) 2009 President's Budget (PB). As a result of the Nunn-McCurdy critical breach determination submitted in September 2007 and the OSD CAIG ICE confirmation of the critical breaches, neither the Procurement aircraft buy quantity profile in the FY09 PB nor the Research Development Test and Evaluation (RDT&E) program is executable within the current approved FY09 PB funding (Nunn-McCurdy certification occurred after submission of the FY 2009 PB). With certification complete, the 716 Aeronautical Systems Group (AESG) will submit a quarterly exception SAR following Milestone C decision for the restructured program.

Quantity	SAR Baseline Dev Est	Current APB Development	Current Estimate
RDT&E	4	3	3
Procurement	122	109	108
Total	126	112	111

Bibliography

Aerospace Systems Center, Air Force Material Command. Memorandum for Policy on Life Cycle Cost Estimates (ASC/CC Policy Memo #05-003) . Wright Patterson AFB OH 14 November 2005.

Air Force Magazine. "FOAs, DRUs, and Auxiliary," Vol. 94, No. 5. May 2011. http://www.airforcemagazine.com/MagazineArchive/Pages/2011/May%202011/0511cover.aspx

AFCAA. *Air Force Cost Risk and Uncertainty Handbook.* Washington: 2007.

AFCAH. *Air Force Cost Analysis Handbook.* Washington: 2008.

Alchain, A.A., (1950). *Reliability of Cost Estimates – Some Evidence.* Santa Monica: RAND.

Arena, M. V., Leonard, R. S., Murray, S. E., & Younossi, O. (1994). *Historical Cost Growth of Completed Weapon System Programs.* RAND Corporation, 1, 1-47.

Arena, M. V., Younossi, O., Galway, L. A., Fox, B., Graser, J. C., Sollinger, J. M., Wu, F., Wong, C., (2006). *Impossible Certainty: Cost Risk Analysis for Air Force Systems.* Santa Monica: RAND.

Bolten, J. G., Leonard, R.S., Arena, M.V., Younossi, O, Sollinger, J.M. (2008). *Sources of Weapon System Cost Growth: analysis of 35 major defense acquisition programs.* Santa Monica: RAND.

Book, S. A., MCR, LLC. "How to Make Your Point Estimate Look Like a Cost-Risk Analysis (so it Can be Used for Decisionmaking). Address to Society of Cost Estimating and Analysis. Manhattan Beach, CA. 15 June 2004.

Cowen, T., Lee, D. "The Usefulness in Inefficient Procurement," Defence Economics, 3: 219-227 (1992).

DAMIR. "Welcome to DAMIR". Excerpt from unpublished article. N. pag. http://www.acq.osd.mil/damir/

Dienemann, P., *Estimating Cost Uncertainty Using Monte Carlo Techniques*. Santa Monica: RAND.

Defense Acquisition University. *Implementation of the Weapon Systems Acqusition Reform Act of 2009.* DOD Directive 09-027. Washington: Under Secretary of Defense for Acquisition, Technology and Logistics. 4 December 2009. https://acc.dau.mil/wsara

Department of the Air Force. *Air Force Cost Analysis Handbook.* Washington: HQ USAF, April 2007.

Department of the Air Force. *U.S. Air Force Cost Risk and Uncertainty Analysis Handbook.* Tecelote Research Inc. April 2007

Department of Defense. *The Defense Acquisition System*. DOD Directive 5000.01. Washington: GPO, 20 November 2007.

Department of Defense. *Operation of the Defense Acquisition System.* DOD Instruction 5000.02. Washington: GPO, 8 December 2008.

Department of Defense. *Defense Acquisition Guidebook*. Washington : U.S. Dept. of Defense, https://dag.dau.mil/Pages/Default.aspx. 2004

Department of Defense. *Department of Defense Budget Fiscal Year 2013*. Washington: Office of the Under Secretary of Defense (Comptroller), February 2012.

Everitt, B.S. *Cambridge Dictionary of Statistics* (2. ed.). Cambridge: Cambridge University Press. 2002.

Fisher, G.H., (1962). *A Discussion of Uncertainty in Cost Analysis (A Lecture for the AFSC Cost Analysis Course).* Santa Monica, RAND.

Flynn, B., Naval Center for Cost Analysis. "Development and Application of CV Benchmarks," Address to Society of Cost Estimating and Analysis. 18 May 2011.

Fox, J.R., Allen, D., Lassman, T.C., Moody, W. S., Shiman, P. L., (2011). *Defense Acquisition Reform, 1960-2009 An Elusive Goal.* Washington. Center of Military History.

Garvey, P., Flynn, B., (2011). *Weapon Systems Acquisition Reform Act (WASARA) and the Enhanced Scenario-Based Method (eSBM) for Cost Risk Analysis.* Williamsburg, NCCA.

Government Accountability Office. *DOD Needs to Take More Action to Address Unrealistic Initial Cost Estimates of Space Systems.* GAO 07-96. Washington. November 2006.

Government Accountability Office. *GAO Cost Estimating and Assessment Guide.* GAO-09-3SP. Washington: March 2009.

Government Accountability Office. *Defense Acquisitions Fundamental Changes Are Needed to Improve Weapon Program Outcomes.* GAO-08-1159T. Washington: 25 September 2008.

Government Accountability Office, *Defense Acquisitions: Improved Management Practices Could Help Minimize Cost Growth in Navy Shipbuilding Programs*. GAO-05-183. (Washington, D.C.: Feb. 28, 2005).

Hald, A. (1952). *Statistical Theory with Engineering Application*. New York: John Wiley & Sons.

Hanks, C.H., Axelband, E.I., Lindsay, S., Rehan Malik, M., Steele, B. D., (2005). *Reexamining Military Acquisition Reform: Are We There Yet?*. Santa Monica: RAND

Hough, P., (1992). *Pitfalls in Calculating Cost Growth from Selected Acquisition Reports.* Santa Monica: RAND.

Hudon, J.L., *Policy on Life Cycle Cost Estimates.* ASC/CC Policy Memo #05-003. Dayton: 14 Nov 2005.

Larsen, R.J., Marx, M.L., *An Introduction to Mathematical Statistics and Its Applications* (3. Ed.). Upper Saddle River NJ: Prentice Hall. 2001.

Lee, D. "The Politics and Pitfalls of Reducing Waste in the Military," Defence Economics, 1:129-139 (1990).

McNaugher, T. L. (1989). *New weapons, old politics: America's military procurement muddle.* Washington, D.C.: Brookings Institution.

McNicol, D.L., (2005). *Cost Growth in Major Weapon Procurement Programs.* Alexandria: Institute for Defense Analyses

Naval Sea Systems Command. (2005). *Cost Estimating Handbook.* Washington: Webster.

Peck, R., Olsen, C., Devore, J., *Introduction to Statistics and Data Analysis.* United States: Duxbury. 2001.

Porter, G., Gladstone, B., Gordon, C., Karvonides, N., Kneece Jr., .R., Mandelbaum, J., and O'Neil, W.D., (2009*). The Major Causes of Cost Growth in Defense Acquisition*. Alexandria: Institute for Defense Analyses.

Sachs, L. *Applied Statistics.* Berlin: Springer-Verlag. 1982.

SAS. "About Us". Excerpt from unpublished article. N. pag. http://www.jmp.com/about/

Schwartz , M. (2013). *Defense Acquisitions: How DOD Acquires Weapon Systems and Recent Efforts to Reform the Process*. Washington: Congressional Research Service.

Younossi, O., Arena, M. V., Leonard, R. S., Roll, C. R., Jain, A., & Sollinger, J. M. (2007). *Is weapon system cost growth increasing? a quantitative assessment of completed and ongoing programs* . Santa Monica: RAND.

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 074-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
21 Mar 2013	Master's Thesis	1 Aug 2011 – 21 Mar 2013

4. TITLE AND SUBTITLE	
Investigation into Risk and Uncertainty: Identifying Coefficient of Variation Benchmarks for Air Force ACAT I Programs	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER**
Carney, Shaun T., Captain, USAF	**5e. TASK NUMBER**
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way, Building 640 WPAFB OH 45433	AFIT-ENV-13-M-05

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
Air Force Life Cycle Management Center Dustin McGlothen 1865 4th St Wright-Patterson AFB, OH 45433 (312)986-5510	LCMC/FZ
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION A: APPROVED FOR PUBLIC RELEASE; DISTRIBUTION IS UNLIMITED

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Previous DoD cost growth studies found typical cost growth in defense acquisitions is around forty-six to sixty percent of the original estimate. The research in this study addresses the identification of risk and uncertainty benchmarks by providing decision makers with coefficient of variation ranges for cost estimates. The study recommends coefficient of variation (CV) ranges for Air Force Acquisition programs, determines if different CV ranges should be used based on platform type, and determines if CV decreases over the course of the program's acquisition lifecycle. The analysis found that the Air Force should enhance the CV review process to ensure cost estimates have CVs between 41-74% during Milestone A, 31-54% during Milestone B, and 23-32% during Milestone C. It is recommended that Selective Acquisition Reports include the CV utilized to develop the current estimate. The analysis found CVs are analogous among platform types. Lastly, the research found that CVs decrease as a program matures through the acquisition lifecycle.

15. SUBJECT TERMS
Risk, Uncertainty, Coefficient of Variation, Cost Estimating

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					Ritschel, Jonathan D., Lt Col, Ph. D, USAF
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)*
U	U	U	UU	104	(937) 255-6565, x 4441 (jonathon.ritschel@afit.edu)